UNIVERZITET U BEOGRADU

FAKULTET ORGANIZACIONIH NAUKA

Olja B. Krčadinac

UTICAJ VARIJACIJA U INTERAKCIJI KORISNIKA NA PERFORMANSE SISTEMA ZA PREPOZNAVANJE GOVORNIKA

doktorska disertacija

Beograd, 2023.

UNIVERSITY OF BELGRADE

FACULTY OF ORGANIZATIONAL SCIENCE

Olja B. Krčadinac

THE EFFECTS OF INTERACTION VARIATIONS ON PERFORMANCE OF THE SPEAKER RECOGNITION SYSTEM

Doctoral Dissertation

Belgrade, 2023

Mentor:

Dr Dušan Starčević, profesor emeritus

Univerzitet u Beogradu, Fakultet organizacionih nauka

Članovi komisije :

Dr Miroslav Minović, redovni profesor

Univerzitet u Beogradu, Fakultet organizacionih nauka

Dr Boško Nikolić, redovni profesor

Univerzitet u Beogradu, Elektrotehnički fakultet

Datum odbrane: _____

Zahvalnica

*Na samom početku bih se zahvalila svom mentoru **prof. dr Dušanu Starčeviću** koji mi je u svakom trenutku pružio nesebičnu podršku. Uvek je bio tu kaže pravu reč u pravo vreme. On mi je predložio temu disertacije i pomogao mi da izaberem ispravan put kojim ću dalje da koračam u naučno-istraživačkom radu. Svaka reč napisana za njega je malo.*

*Posebno sam zahvalna mojim roditeljima **Jeleni i Branku** koji su mi uvek ulivali nadu i poverenje da će se sve na kraju dobro završiti. Zahvaljujem se i mom bratu **Borisu** koji je strpljivo čitao rad kad god je to bilo potrebno i bodrio sestru sa drugog kraja planete.*

*Veliku zahvalnost dugujem svom suprugu **Igoru** koji mi je bio najčvršći oslonac pri pisanju doktorske disertacije i koji je svaki put staloženo saslušao moje nove ideje. Najmiliju zahvalnost posvećujem svojim sinovima **Kosti** i **Pavlu** koji su mi pomagali samim pogledom na njih, jer su mi pružili najveću motivaciju za dalji uspeh.*

Poslednju zahvalnost, ali ne i najmanje važnu dugujem svojim kolegama sa Fakulteta organizacionih nauka, kao i svojim radnim kolegama koji su me podržavali tokom izrade disertacije.

Uticaj varijacija u interakciji korisnika na performanse sistema za prepoznavanje govornika

Sažetak

Ključne reči: biometrija, prepoznavanje govornika, rešenja s otvorenim kodom, menadžment identiteta

Naučna oblast: Informacioni sistemi i tehnologije

Uža naučna oblast: Biometrijski informacioni sistemi

The effects of interaction variations on performance of the speaker recognition system

Abstract

Keywords: biometrics, speaker recognition, open source solutions, identity management

Scientific area: Information systems and technologies

Specific scientific area: Biometric information system

SADRŽAJ

1. UVOD

Opšta primena Internet tehnologija u procesu globalizacije omogućava korisnicima da pristupe željenim resursima nekog informacionog sistema praktično sa bilo kojeg mesta, sa bilo kojim uređajem i u bilo koje vreme. Međutim, takva ponuđena tehnološka i tehnička pogodnost otvara velik prostor za zloupotrebe. Štete koje tako nastaju mogu biti, na primer, ekonomske, političke, pravne, bezbedonosne, etičke, kulturne prirode. Kao najčešće protivpravne radnje se javljaju slučajevi prevara i krađa na Internetu, nedozvoljenog praćenja, spama ili onemogućivanje korišćenja neke usluge na Internetu u nekom vremenskom periodu.

Navedene pretnje sigurnom načinu korišćenja distribuiranih računarskih resursa zahteva koncipiranje, razvoj i implementaciju odgovarajućih odbrambenih mera, na sistemskom i aplikativnom nivou rada. Posebno osetljiva tačka je pristup korisnika informacionom sistemu. Od početnog jednostavnog načina predstavljanja korisnika računarskom sistemu posredstvom korisničkog imena i lozinke, postupci kontrolisanog pristupa sistemu su se znatno unapredili. Danas su razvijeni i u upotrebi složeni sistemi poznatim pod opštim nazivom Menadžment identiteta odnosno sistem za upravljanje identitetom koji se odnosi na informacioni sistem ili na skup tehnologija koje se mogu koristiti za upravljanje identitetom preduzeća ili više mreža (Cao & Yang, 2010). Svi takvi postupci za pristup sistemu menadžmenta identiteta mogu se prema Nacionalnom institutu za standarde i tehnologiju Sjedinjenih Američkih država (NIST) klasifikovati prema tome šta korisnik u procesu pristupa pruža na uvid računarskom sistemu kao "nešto šta samo korisnik treba da zna, nešto šta samo korisnik treba da poseduje ili nešto što je deo samog korisnika" (Liu, i drugi, 2011).

Posebno je interesantna treća grupa rešenja koja se odnosi na sintagmu "nešto šta je deo samog korisnika" i koja je u naučnoj literaturi poznata pod nazivom biometrijske tehnologije za pristup računarskom sistemu. Podaci koji se odnose na građu i/ili prirodu ponašanja samog korisnika se posredstvom biometrijskih tehnologija zahvataju i saopštavaju u računarsko čitljivom obliku bezbedonosnom mehanizmu informacionog sistema. Od posebnog interesa za biometrijske tehnologije su podaci o anatomskim i fiziološkim karakteristikama našeg tela. Ovi podaci se zahvataju posebnim uređajima, najčešće senzorskim kao neinvazivnim, i prevode u računarski čitljiv oblik kao *biometrijske* karakteristike. Tu pripadaju, pre svega, otisci prstiju, otisci dlana,

slika mrežnjače, slika šarenice, slika lica, DNK uzorak. Pokušaj transformacije obeležja eksperimentisan je metodologijom kodiranja zasnovanom na DNK gde možemo videti tehniku koja je konstruisana iz biometrijskih podataka za generisanje ekstrakcije karakteristika (Jacob, i drugi, 2021).

Druga grupa biometrijskih tehnologija se bavi prirodom našeg ponašanja u interakciji sa okruženjem (Maghsoudi & Tappert, 2016). Navedimo, kao primere, karakteristike svojeručnog potpisa, način i dinamiku hoda, geste i mimiku lica.

Neke biometrijske karakteristike nije jednostavno klasifikovati, jer imaju više klasifikacionih svojstava. Ljudski govor pripada takvoj grupi biometrijskih karakteristika. Podaci koji se odnose na prevođenje govora u računarski čitljiv oblik su posledica urođenih i stečenih biometrijskih svojstava. Pored statičkih prostornih obeležja definisanih anatomskom strukturom govornih organa, na dinamičke karakteristike govora utiču stečena iskustva u govornoj komunikaciji, a koja prepoznajemo kroz način izgovora glasova, akcentovanje, ritam govora, intonaciju, korišćeni rečnik. Specifično, za govor je eksplicitno izražena vremenska komponenta zvučnog zapisa.

Mogućnost bezbedne upotrebe nekog aplikativnog rešenja praktično sa bilo kojeg pristupnog uređaja, sa bilo kojeg mesta i u bilo koje vreme, sa korisnikom koji može biti i funkcionalno hendikepiran za korišćenje tradicionalnih korisničkih formi, stavlja projektanta sistema pred velike izazove u procesu razvoja programskih rešenja namenjenih korišćenju posredstvom Interneta. Prepoznavanje govornika, tipično u procesu verifikacije, na osnovu njegovog glasa postaje važna karika u dizajniranju i implementaciji korisničkih interfejsa, orijentisanih prvenstveno na fizičke mogućnosti i ograničenja korisnika u pristupu informacionom sistemu. Arhitektura savremenih aplikativnih programskih sistema promoviše ključnu važnost interakcije korisnika i sistema pri čemu je korisnički interfejs pre svega orijentisan na karakteristike korisnika (*human-centric interfaces*).

Problem koji se razmatra u doktorskoj disertaciji odnosi se na ispitivanje mogućnosti i ograničenja koja poseduju raspoloživa i dostupna programska rešenja sa otvorenim programskim kodom za prepoznavanje osoba na osnovu glasa, a koja u realnim uslovima rada se menjaju kao rezultat varijacija u interakciji korisnika sa računarskim sistemom, i koja projektant aplikativnih rešenja za

rad na Internetu treba imati u vidu, prilikom korišćenja govora kao biometrijske karakteristike u sistemima menadžmenta identiteta.

Predmet rada ove doktorske disertacije je sistem za prepoznavanje osoba na osnovu glasa realizovan dostupnim rešenjima iz domena otvorenog programskog koda. Istražen je uticaj pojave razlika u eksploatacionom režimu rada sistema za prepoznavanja govornika u odnosu na režim rada u razvojnoj fazi (obuka sistema) na eksploatacione performanse sistema, a koje nastupaju kao rezultat varijacija u interakciji korisnika sa predmetnim sistemom. Analizirale su se razlike koje mogu nastupiti u više domena interakcije: nepodudarnost u akvizacionim uređajima za snimanje glasa u procesima obuke i korišćenja sistema, nepodudarnost u radnom okruženju u kojem se odvija akvizicija zvučnog zapisa u toku obuke i korišćenja sistema, nepodudarnost u dužini zvučnog zapisa iz procesa obuke i eksploatacionog režima rada, i nepodudarnost u jeziku koji je korisnik koristio tokom obuke, odnosno tokom korišćenja sistema. Za pristupne uređaje u procesu istraživanja koristio se konvencionalni stoni računar sa priključenim mikrofonom, laptop računar sa integrisanim mikrofonom i smart telefon, imajući u vidu sve češće korišćenje mrežnih aplikacija u režimu mobilnog korisnika. Kao moguće radno okruženje koristio se prostor Laboratorije za multimedijalne komunikacije Fakulteta organizacionih nauka. Govorni zapisi ispitanika razvrstani su u tri skupa po dužini trajanja, i u tri skupa po semantičkoj interpretaciji sadržaja zapisa. Takođe, analizirao se i efekat korišćenja dva različita jezika, srpskog i engleskog, u procesu interakcije na performanse sistema za prepoznavanje osoba na osnovu glasa.

Osnovni cilj istraživanja bilo je nalaženje odgovora na pitanje koliko uobičajene varijacije u interakciji korisnika sa sistemom za prepoznavanje korisnika na osnovu glasa mogu uticati na degradaciju funkcionalnih performansi takvih sistema, kao gradivnih elemenata menadžmenta identiteta u mrežnim aplikacijama u realnom vremenu.

Specifični podciljevi koji su ostvareni tokom istraživanja su:

- Sistematizacija i prezentacija osnovnih metodoloških postupaka za prepoznavanje osoba na osnovu glasa;
- Analiza funkcionalnih mogućnosti dostupnih softverskih rešenja sa otvorenim programskim kodom za razvoj sistema za prepoznavanje osobe na osnovu glasa;
- Razvoj odgovarajuće metodologije vrednovanja uticaja varijacija u načinu interakcije čoveka i sistema za prepoznavanje govornika na performanse takvih sistema;

- Formiranje radnog okruženja sa akvizicionim uređajima za snimanje zvučnih zapisa: računar sa priključenim mikrofonom, tablet i smart telefon;
- Formiranje više lista tekstualnih zapisa različite dužine (kratki, srednji i duži tekst) na srpskom i engleskom jeziku. Svaka lista tekstova je specifičnog karaktera. Tekstovi iz prve liste su iz računarskog rečnika, češće korišćenog u komunikaciji posredstvom Interneta. Tekstovi iz druge liste su iz domena bankarskih transakcija, dok tekstovi iz treće liste obuhvataju svakodnevne aktivnosti ljudi;
- Izbor i obuka dve grupe ispitanika za snimanje zvučnih zapisa. Na osnovu delova prve grupe zvučnih zapisa ispitanika u procesu obuke se formirana je baza modela ispitanika kao govornika. Na osnovu drugog dela zvučnog zapisa iz prve grupe u eksploatacionom režimu rada izmerene su funkcionalne performanse sistema u različitim uslovima interakcije. Zvučni zapisi iz druge grupe su se koristili za ispitivanje performansi sistema u radu sa uljezima;
- Istovremeno snimanje zvučnih zapisa pročitanih tekstova posredstvom tri akvizaciona uređaja u studijskom i laboratorijskom prostoru u više sesija;
- Priprema snimljenih zvučnih zapisa (pretprocesiranje) za dalju obradu imajući u vidu NIST Speaker Recognition Evaluation (NIST SRE) preporuke;
- Na osnovu dostupnih rešenja u otvorenom programskom kodu, formirano je više instanci eksperimentalnih laboratorijskih sistema za prepoznavanje osobe na osnovu glasa;
- Specifična obuka više instanci laboratorijskih sistema za prepoznavanje osobe koristeći formirane liste zvučnih zapisa;
- Pripremljena su odgovarajuća ispitna scenarija, koja opisuju varijacije u interakciji korisnika i sistema, i na osnovu njih za svaku instancu laboratorijskog eksperimentalnog sistema za prepoznavanje govornika su formirane različite ispitne konfiguracije elemenata sistema;
- Na osnovu formiranih ispitnih konfiguracija elemenata sistema na svakoj instanci laboratorijskog sistema za prepoznavanje osoba na osnovu glasa izvršeno je eksperimentalno ispitivanje uticaja izloženih varijacija na funkcionalne i nefunkcionalne performanse posmatranih sistema i prikupljanje odgovarajućih podataka;

- Na osnovu prikupljenih podataka o ponašanju laboratorijskih sistema za prepoznavanje osoba na osnovu glasa izvršena je odgovarajuća statistička obrada i na pogodan način su prikazani rezultati obrade;
- Interpretirani su dobijeni rezultati u pogledu robustnosti ponaosob za svaku instancu laboratorijskih eksperimentalnih sistema za prepoznavanje osoba na osnovu glasa;
- Evaluirana su i rangirana dostupna rešenja za prepoznavanje osoba na osnovu glasa sa otvorenim kodom u smislu pogodnosti korišćenja takvih rešenja u sistemima menadžmenta identiteta.

U drugom poglavlju prikazan je opis biometrijskih sistema i karakteristika. Posebna pažnja posvećena je načinu rada biometrijskih sistema i tehnikama koje se koriste za potrebe evaluacije. Takođe, deo ovog poglavlja je menadžment identiteta u kojem postoji mogućnost primene biometrije. Biometrijske tehnologije predstavljaju prirodno oruđe poboljšanja privatnosti u sistemima za upravljanje identitetom.

Treće poglavlje sadrži pregled i klasifikaciju postojećih pristupa u oblasti sistema za prepoznavanje govornika, kao i izgled opšte arhitekture sistema za prepoznavanje govornika. U ovom poglavlju se nalazi kritički osvrt na postojeće pristupe što predstavlja bitan segment rada gde svaka biometrijska tehnologija koristi različite biometrijske senzore i različite algoritme za uparivanje biometrijskih podataka.

Četvrto poglavlje se bavi klasifikacijom i pregledom četiri odabrana programska rešenja za prepoznavanje govornika sa otvorenim programskim kodom. Ispitane su mogućnosti svakog od predstavljenih softvera. Dat je sumarni prikaz funkcionalnosti sva četiri programska rešenja (SPEAR, MARF, ALIZE, HTK).

U petom poglavlju predstavljena je metodologija vrednovanja uticaja varijacija u načinu interakcije korisnika sa sistemom na performanse sistema za prepoznavanje govornika. U ovom poglavlju je opisan rezultat istraživanja na kompleksnom zadatku vrednovanja podataka, u kome je osmišljen i primenjen algoritam za statističko vrednovanje podataka na bazi relacija između merenih veličina.

Šesto, centralno, poglavlje posvećeno je studiji slučaja istraživanja uticaja varijacija u interakciji korisnika sa sistemom za prepoznavanje govornika na performanse takvih sistema kada se koriste dostupni elementi rešenja sa otvorenim programskim kodom. Opisani su postupci koji prethode

formiranju baze zvučnih zapisa, postupci formiranja baze zvučnih zapisa, postupci pretprocesiranja baze zvučnih zapisa i postupci korišćenja baze zvučnih zapisa u procesu vrednovanja odabranih programskih rešenja sa otvorenim kodom, odnosno algoritama koje sadrže, prilikom implementacije osnovnih funkcija menadžmenta identiteta u distribuiranim informacionim sistemima. Definisan je model podataka koji će služiti kao strukturna osnova sistema. U ovom poglavlju prikazana je empirijska provera predloženog postupka vrednovanja uticaja varijacija u interakciji korisnika i sistema za sva četiri odabrana programska rešenja sa otvorenim kodom. Eksperimentalnim laboratorijskim sistemom ispitani su uticaji vremenske dužine uzorka (dužina reči/rečenica), (ne)podudaranje uređaja, kao i jezička podudarnost (srpski, engleski) na krajnju preciznost u prepoznavanju govornika. Dati su i rezultati statističke obrade prikupljenih podataka za svako razmatrano programsko rešenje sa otvorenim kodom.Sedmo poglavlje se odnosi na evaluaciju međusobnim upoređivanjem rezultata dobijenih eksperimentalnim laboratorijskim sistemom za prepoznavanje osoba na osnovu glasa i izvršeno je rangiranje sva četiri ispitivana rešenja sa otvorenim programskim kodom.

Poslednje, osmo poglavlje donosi zaključke o dobijenim rezultatima. Naročita pažnja usmerena je ka ostvarenim naučnim i stručnim doprinosima. Predložene su različite upotrebe biometrijskog sistema i utvrđeni su budući koncepti razvoja i istraživanja.

2. MENADŽMENT IDENTITETA I BIOMETRIJSKI SISTEMI IDENTIFIKACIJE

2.1. MENADŽMENT IDENTITETA

Na najosnovnijem nivou, menadžment identiteta podrazumeva definisanje šta korisnici mogu da rade na mreži sa specifičnim uređajima i pod određenim okolnostima. Danas, mnogi bezbedonosni sistemi stavljaju naglasak na menadžment identiteta u korporativnom sistemu putem mobilnih uređaja gde postoji mogućnost primene biometrije. U okviru kompanija, menadžment identiteta se koristi za povećanje bezbednosti i produktivnosti, smanjujući troškove i prekomerni napor u radu.

Menadžment identiteta obuhvata procese i pravila upravljanja životnim ciklusom identiteta za određeni domen (ISO/IEC, 2011). To je funkcija koja je od suštinskog značaja za menadžment identiteta, odnosno sigurna veza izmedju identifikatora (vrednost koja nedvosmisleno razlikuje jednog korisnika od drugog u određenom domenu) i atrubuta (sertifikati; zahtevi; prava; svojstva korisnika kao što su npr. ime, starost, kreditni rejting itd.). Prvi koraci koje bi bilo potrebno preduzeti su prilagođavanje upotrebe DLT-a (tehnologije distribuiranog registra) za uspostavljanje sigurnog i decentralizovanog vezivanja identifikatora i atrubuta u Namecoin šemi (najtrajnija programska forma Bitcoin-a) (Dunphy, 2018). Isto tako, ako se uključi biometrija u celo istraživanje menadžmenta identiteta dobijamo proveru identiteta jedne osobe sa drugom, gde se uporedjuje zapis iz npr. prve baze podataka sa zapisom iz druge baze podataka, a gde baze rade nezavisno jedna od druge. Takav primer objašnjava KJ Hanna (Hanna, 2018), odnosno svaki zapis sadrži i biometrijski zapis i odgovarajući identifikator jedinstven za sve baze podataka. Ako se prvi i drugi zapis od iste osobe, prvi i drugi zapis su povezani sa jedinstvenim identifikatorom. Na taj način obezbeđuje se pristup drugoj bazi podataka. Blokčein (eng. *block* - blok i eng. *chain* - lanac) datira od devedesetih godina (Arnold, 1991). Blokčein je registar svih transakcija koje su se ikad desile u bitkoinovom sistemu (Xu, Weber, & Staples, 2017). Svaki program u kojem se izvršava neki oblik plaćanja mora da poseduje podatke o transakcijama koji se formiraju se u bazu podataku. Značajna novina kod blokčeina je metod na koji se vrši slanje i skladištenje informacija o transkakcijama. Blokovi sadrže transakcije, a oni dalje idu u lanac. Da bi se vezali blokovi mora da dođe do kriprtografije, odnosno HASH funkcije što znači da ne bi smelo da se koriguje sadržaj blokova kako ne bi došlo do izmene ostalih blokova. To je važna stavka kako ne bi došlo do

promene podataka u blokčeinu (Matanović, 2018). On se primenjuje na metode kao što su pametni grad, lanac snabdevanja, čuvanje i deljenje medicinskih podataka, itd. Mnogi radovi su usmereni ka poboljšanju performansi i sigurnosti jednog takvog sistema. Međutim, postoji nedostatak mehanizma identitifikatora u stvarnom svetu sa sistemom blokčeina. Kako bi se ispravno radio taj mehanizam, potreban je BlockID, novi okvir za upravljanje identitetom ljudi koji koriste biometrijsku autentifikaciju i pouzdanu računarsku tehnologiju (Gao Z. X., 2018, June). Zanimljivo je spomenuti rad koji se bavi sistemom upravljanja voznim parkom u poljoprivrednoj industriji (Achillas, 2019), gde se govor može iskoristiti u sferi menadžementa identitet za autentifikaciju i prižanje tokena za pristup korisnicima. Arhitektura jednog takvog sistema nezavisna je od prodavca tog sistema, interakcija je vođena glasom, te se funkcionalnost i podrška planiranja rada prenosi kroz aplikaciju koja organizuje poljoprivrednu mehanizaciju. Takav alat mogu koristiti krajnji korisnici poput poljoprivrednika, izvođača radova, udruženja farmi, skladišta, firmi za preradu poljoprivrednih proizvoda i sl.

COVID-19 je napravio prostora za nova istraživanja, kako u zdravstvu, tako i u drugim sferama. Jedna od njih je u oblasti integracije menažmenta identiteta i zdravstva. U toku pandemije predstavljen je decentralizovani sistem za upravljanje identitetom zasnovan na blokčejnu koji omogućava pacijentima i zdravstvenim radnicima da se identifikuju i autentifikuju na transparentan i siguran način u različitim domenima e-zdravstva. Pacijenti i pružaoci zdravstvenih usluga se jedinstveno identifikuju po njihovim zdravstvenim identifikatorima (healthID) (Javed, i drugi, 2021) .

2.2. BIOMETRIJSKI SISTEMI IDENTIFIKACIJE

Kao prvo, potrebno je spomenuti latinsku reč *identificare* od koje i potiče reč identifikacija koja znači utvrđivanje istovetnosti, odnosno povezanost određenog podatka o ličnosti sa njom samom (Paunović, Starčević, & Nešić, 2013). Zatim, napominjemo da je biometrija disciplina koja je razvojem nauke, a pre svega digitalno-informacione tehnologije, počela da dobija sve značajnije mesto u različitim područjima društvenog života, ljudskog rada i interesa.

Kada je reč o karakteristici, onda ona označava merljivu biološku (fiziološku) karakteristiku ili karakteristiku (bihejvioristička) ponašanja, koja se može koristiti za automatsko prepoznavanje (Jain, Hong, & Pankanti, 2000).

Ako govorimo o slučaju kada se koristi za opis procesa, onda predstavlja automatizovan metod za prepoznavanje osoba, koji se zasniva na biološkim karakteristikama ili karakteristikama ponašanja.

Ideja o biometriji dolazi iz starog Egipta (Gao & Xiang, 2010) gde su ljudi tvrdili da oni koji su na osnovu telesnih mera i karakteristika. Postoje dve funkcije koje se baziraju na biometrijskim sistemima kod modernih računara. Jedna se odnosi na identifikaciju ("ko je osoba?") kod koje se porede izmerene biometrijske karaktetistike sa bazom poređenih zapisa. Druga predstavlja proveru ("da li je to osoba koja kaže da je?"). Tu se vrši upoređivanje 1:1 (s jedne strane izmerene identifikacione odlike i sa druge karakteristike koje važe za određenu osobu).

Jedan od najosnovnijih primera karakteristika, koje se koriste za prepoznavanje nekog čoveka, od strane drugih ljudi, je lice. Od početka civilizacije, ljudi koriste lice za prepoznavanje osoba. Takođe, prepoznavanje ciljnih osoba vidljivo je i u prepoznavanju govora, kao i preko ostalih karakteristika svakodnevnog ponašanja. Integracija informacija sa lica i glasa igra centralnu ulogu u našim društvenim interakcijama. Uglavnom se proučava u kontekstu audiovizuelne percepcije govora: integracija afektivnih ili identitetskih informacija dobija relativno malo naučne pažnje. U istraživanjima se razmatraju bihevioralne i neuroimaging studije integracije lica i glasa u kontekstu percepcije osobe. Uočeni su jasni dokazi o smetnji između informacija o licu i glasu tokom prepoznavanja afekta ili obrade identiteta (Campanella & Belin, 2007).

Tokom istorije civilizacije, ostale karakteristike su mnogo formalnije, a to su otisci šaka u pećinama (smatra se da ih je praistorijski čovek ostavio pre 40000 godina) (Wang, Ge, Snow, Mitra, & Giles, 2010). Zatim, kineski trgovci su poslovne transakcije potvrđivali otiscima prstiju (Babaei, Molalapata, & Pandor, 2012), kao i trgovci u Vavilonu oko 500. godine p.n.e. (Benson, 2018) itd.

Kako su se počeli razvijati gradovi u devetnaestom veku, dolazi do sve veće pokretljivosti populacije, razvija se industrija, te se javlja potreba za pouzdanom identifikacijom ljudi.

Prvi pokušaj je bio antropometrija, odnosno Bertillon-ov sistem merenja različitih karakteristika tela (Kleinberg, Vanezis, & Burton, 2007). Drugi pokušaj je bio formalno korišćenje otisaka prstiju od strane policije (Kaushal & Kaushal, 2011).
Prvi robustni sistem za indeksiranje otisaka prstiju je razvijen u Indiji za generalnog inspektora policije u Bengalu, Edvarda Henrija – Henrijev sistem (Cowger, 2020). On se još uvek koristi za klasifikaciju otisaka prstiju (Souza, Medeiros, Holanda, Rego, & Reboucas Filho, 2021), kao i neke

njegove varijacije (Tait, ,2021).

Kako su se počeli razvijati računarski sistemi, tako i biometrijski sistemi počinju dobijati na značaju u drugoj polovini dvadesetog veka. Sve većim prodorom ove oblasti, biometrijske aplikacije postaju svakodnevnica (Cavoukian, 1999).

Biometrijski sistemi su tehnički sistemi koje kao podatke koriste biometrijske karakteristike ljudi. Ovi sistemi mogu da rade u više načina rada, od kojih su najinteresantniji režim upisa biometrijskih podataka radi formiranja baze podataka (Varchol & Levicky, 2007), te režim identifikacije i verifikacije osobe na osnovu upitnih podataka.

2.2.1. UPIS

Faza upisa se odnosi na unošenje novog subjekta u bazu podataka na osnovu biometrijskih parametara. Ovaj proces se odvija ukoliko subjekat nije prethodno bio upisan u bazu. Označava najbitniji proces koji se često spominje u referencama (Rane, 2014) gde važi za nužnu stavku u fazi identifikacije i verifikacije (Femila & Irudhayaraj, 2011). rezultat uspešnosti jednog biometrijskog sistema zavisi od akvizicije biometrijskih podataka koja je u suštini vrlo kompleksan postupak (Milovanović, 2013). Izbor senzora predstavlja bitan deo sistema. Zatim, uslovi u kome se uzorak snima, algoritmi koji na osnovu sirovog redudantnog akvizicionog podatka izvode odgovarajuću biometrijsku karakteristiku, formata zapisa u kome se izvedeni biometrijski podaci - šabloni čuvaju i načina organizacije podataka na trajnom medijumu (datoteke ili baze podataka). Kasnije performanse sistema mogu puno zavisiti od formata u kom se podaci čuvaju zbog toga što neki formati nude kvalitet podataka na visokom nivou. Uporedo sa tim se produžava neophodno vreme procesiranja. Potrebno je adekvatno organizovati postupak upisa kako bi se dobili što bolji rezultati.

Kad se unesu biometrijski podaci u bazu podataka, u procesu operativnog korišćenja sistema sledi proces utvrđivanja identiteta osobe koja koristi biometrijski sistem. Na raspolaganju su dva osnovna načina rada biometrijskog sistema: metod identifikacije i metod verifikacije. I u procesu identifikacije i verifikacije vredi podela na četiri faze ovog procesa: uzimanje biometrijskog uzorka, ekstrakcija odgovarajućeg obeležja, upoređivanje sa zapisom iz baze biometrijskih podataka i interpretacija rezultata poređenja.

2.2.2. VERIFIKACIJA

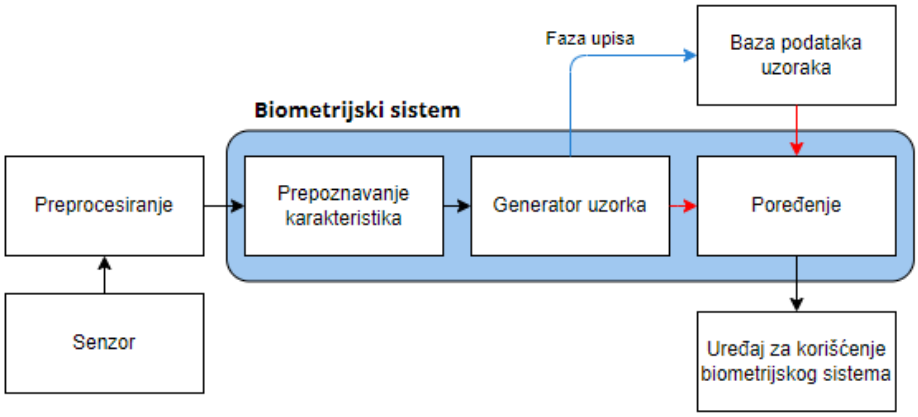
Nakon što se izvrši proces registracije, odnosno formiranja baze biometrijskih podataka, u režimu verifikacije, odnosno provere tvrđenog identiteta, korisnik prvo unosi u sistem tvrdnju o svom identitetu. Zatim, sistem uzima odgovarajući ispitni biometrijski uzorak korisnika i na osnovu njega i određenog algoritma generiše ispitni šablon. Nakon toga, po principu 1:1 vrši se upoređivanje ispitnog šablona sa ranije zapisanim referentnim šablonom istog korisnika.

Verifikacioni sistemi mogu da sadrže milione registrovanih šablona. Međutim, činjenica da koristi princip rada 1:1 predstavlja osnovnu prednost u odnosu na identifikacione biometrijske sisteme, koji rade po principu 1:N, odnosno koji za jedan upitni šablon pretražuju kompletnu bazu podataka (Bolle, Ratha, & Pankanti, 2004). Identifikacioni sistemi su znatno sporiji kada je u pitanju veliki broj referentnih šablona u bazi, jer se kod verifikacionih podudarnost sa referentnim šablonom jedne osobe može utvrditi u deliću sekunde. Ali, u radu verifikacionih sistema postoji i potencijalno značajan nedostatak jer se od korisnika prilikom unošenja podataka o tvrđenom identitetu u sistem može zahtevati lozinka ili kartica, pri čemu može doći do zaboravljanja lozinke ili gubitka kartice.

Aplikacija verifikacije se koristi u slučajevima u kojima se želi potvrda identiteta zaposlenika za pristup zaštićenim područjima ili samim računarima.

Na slici 1 je prikazan biometrijski sistem koji se sastoji od:

- senzora
- pretprocesiranja
- dela za izvlačenje karakteristike
- generatora uzorka
- baze podataka uzoraka
- poređenja
- korišćenje biometrijskog sistema (uređaj)



Slika 1 Proces verifikacije korisnika

Faza upisa je prva faza kada korisnik prvi put koristi sistem. U toj fazi se snimaju podaci, te upisuju u bazu. U sledećim korišćenjima upotrebama programa, podaci koji su snimaju pri ulazu se upoređuju sa podacima koji su već zapamćeni u bazi, te se na taj način vrši provera identiteta. Prvi modul predstavlja interfejs koji u stvari prikuplja neophodne podatke Drugi modul se odnosi na deo za preprocesiranje gde se uklanja signal smetnje iz ulaznog signala, odnosno da pročisti, pojača, normalizuje ulazne podatke i pripremi ih za dalju obradu. U preprocesiranim podacima treba prepoznati podatke od interesa za izdvajanje zahtevanih biometrijskih karakteristika i izvući ih na optimalan način u pogledu korišćenih sistemskih resursa. Navedene karakteristike služe za formiranje uzorka. Ako dođe do faze upisa, uzorak se struktuira i čuva. Kada dođe do faze poređenja ispitni uzorak se šalje modulu za poređenje gde se poredi sa uzorcima u bazi.

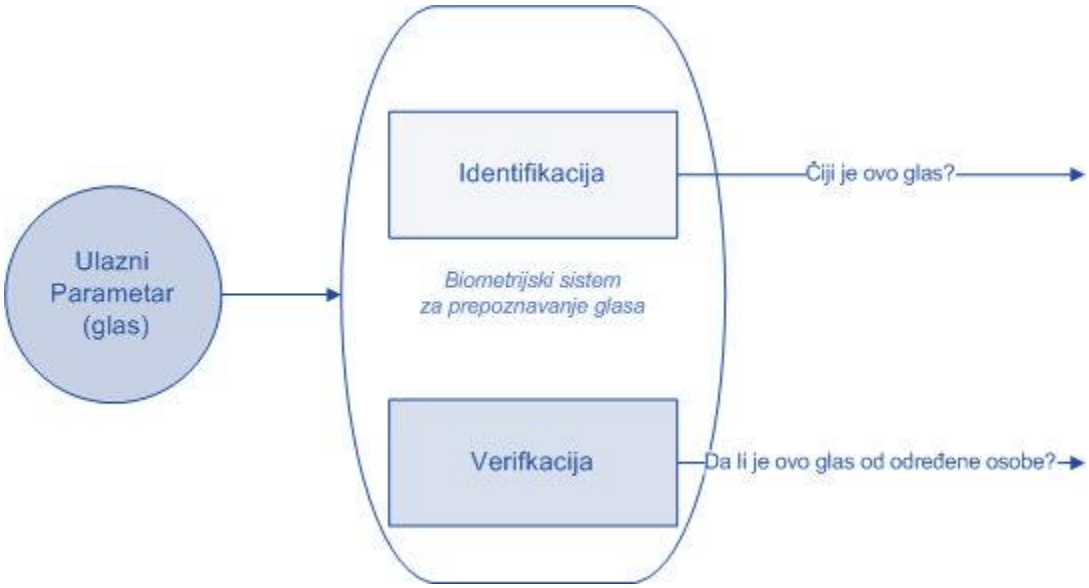
2.2.3. IDENTIFIKACIJA

Tok procesa identifikacije je sledeći: kad se obavi proces registracije biometrijskih podataka u sistem, u procesu operativne identifikacije biometrijski sistem od ispitne osobe uzima biometrijski uzorak i generiše ispitni šablon. Sistem treba da odgovori na pitanje "Ko je ova osoba?", te počinje upoređivati ispitni šablon sa svim pohranjenim podacima u bazi šablona, odnosno vrši upoređivanje na principu 1:N.

Imamo dva tipa sistema za identifikaciju: pozitivni i negativni. Pozitivni identifikacioni sistemi se odnose na princip nalaženja ispitnog uzorka osobe kao registrovanog biometrijskog podatka u bazi podataka sistema, te se traži podudaranje rezultata. Negativni identifikacioni sistemi pokušavaju da potvrde nepostojanje registrovanog biometrijskog referentnog šablona osobe koja se želi

identifikovati u bazi podataka, pa se gleda gde se rezultati ne preklapaju. Kako se vrši upoređivanje biometrijskih podataka sa poznatim obrascima iz baze, obezbeđuje se nepostojanje sumnjivih dokumenata. Znači, osoba ne može imati više ličnih identiteta.

Na slici 2 je predstavljen prikaz razlike između identifikacije i verifikacije u biometrijskim sistemima koji koriste glas, odnosno govor, kao biometrijsku karakteristiku. Identifikacija govornika je proces određivanja od kog od registrovanih govornika potiče dati ispitni glasovni iskaz. Verifikacija govornika je proces prihvatanja ili odbijanja tvrđenog identiteta govornika. Ovaj primer se odnosi samo na razliku u semantici izlaznih vrednosti. U nekim slučajevima, verifikacija govornika će uslediti odmah nakon identifikacije kako bi se potvrdio rezultat identifikacije (Park & Hazen, 2002).



Slika 2 Razlika između identifikacije i verifikacije

2.3. SISTEMI IDENTIFIKACIJE NA OSNOVU GLASA

Prva istraživanja iz oblasti prepoznavanja ljudskog glasa počela su tridesetih godina XX veka (Hoffmann, 2010). U početku su se istraživanja vršila nad malim bazama koje su sadržale čist, kontrolisan govor, da bi se postepeno došlo do velikih, realističnih baza podataka gde je govor apsolutno proizvoljan (Rabiner & Juang, 1993). Do 2001. godine već su bile poznate skoro sve tehnike i algoritmi koji se koriste u današnjem procesu prepoznavanja govornika. Od tog momenta,

sva istraživanja u ovoj oblasti su vezana za eksperimentisanje sa postojećim saznanjima i svode se na fino podešavanje već poznatih algoritama (Bell, i drugi, 2020).

Tokom vremena, razvila su se različita mišljenja i teorije o tome koje algoritme i na koji način treba iskoristiti u sistemu za prepoznavanje glasa kako bi se dobili najbolji rezultati. Nakon mnogih eksperimenata i istraživanja, došlo se do saznanja da se korišćenjem algoritama za upoređivanje šablona dobijaju najbolji rezultati kada je u pitanju tekstualno zavisno prepoznavanje (Gaikwad, Gawali, & Yannawar, 2010) . Sa druge strane, skriveni modeli Markova su se pokazali kao odličan izbor za formiranje modela govornika kod tekstualno nezavisnih sistema. U istraživanjima se vrlo često spominju i Gausovi mešoviti modeli, čije je korišćenje gotovo nezaobilazno u procesu izdvajanja spektralnih karakteristika kod sistema sa tekstualno nezavisnim prepoznavanjem.

Prepoznavanje govora i prepoznavanje govornika su bliske, ali ipak različite naučne discipline, pri čemu im je istraživanje glasa govornika zajednička osnova. Pojavom personalnih računara osamdesetih godina prošlog veka, poraslo je interesovanje za prepoznavanje govora u cilju automatskog prevođenja govora u tekst, kao zamena za daktilografske poslove. Prepoznavanje govornika na osnovu glasa najpre se javilo u sistemima klasične telefonije na zahtev sistema nacionalne bezbednosti, a interes se ubrzao razvojem mobilne telefonije i Interneta. Prepoznavanje govora je problem veće složenosti od problema prepoznavanje govornika.

Ako se ima u vidu da su algoritmi za prepoznavanje govora ključni element rešenja za prepoznavanje govornika, u prikazu razvoja odgovarajućih tehnologija možemo poći od tehnologija za prepoznavanja govora, a da se ne gubi opštost prikaza problema.

Među prvim istraživanjima već se spominje pet značajnih tehnologija za prepoznavanje govora govornika, a koje su i dalje dostupne ili se još i poboljšavaju (Huang, 1993):

1. **Prepoznavanje govora koje zavisi od obuke sistema** – Ova vrsta tehnologije uključuje prethodni trening sistema kako bi se u eksploataciji sistema prepoznao govorni uzorak. Takvi sistemi bi trebalo da imaju fond od 30000 do 120000 reči. Najbolji rezultati se postižu u slučaju kada ga koristi korisnik koji je već prošao trening. Trening proces, odnosno proces obuke, obuhvata snimanje ključnih reči/rečenica koje se kasnije koriste u test fazi

(prepoznavanje). Proces obuke uključuje unos velikih količina podataka u algoritam i omogućavanje mu da uči iz tih podataka i identifikuje obrasce. Ovi sistemi su obično lakši za razvoj i jeftiniji za uvođenje u rad. Sistem je obučen da razume izgovore, fleksije i akcente nekog korisnika i može da radi mnogo efikasnije i preciznije. Zahteva od korisnika da učestvuju u sesijama obuke koje „uče“ računar da prepoznaje glas korisnika. Računar tada pravi glasovni profil koji odgovara potrebnoj obuci. Tehnologija koja zavisi od glasa govornika, tako što je sistem "obučen" kroz ponavljanje da prepozna određeni rečnik reči i ne prihvati nikakvu zamenu, prilično je dobro razvijena. Ova tehnologija je uglavnom zasnovana na šablonskom, ili akustičnom, predstavljanju govora. Korisnici „obučavaju“ sistem u svojim posebnim glasovnim obrascima izgovaranjem reči ili korišćenjem glasovnih uzoraka koji će morati da budu prepoznati. Ovi „glasovni uzorci” ili šabloni se zatim čuvaju u sistemu (Najka, 2018). Kada sistem radi, glasovni uzorci se upoređuju sa izgovorenom komandom korisnika. Ako se otisak glasa i izgovorena reč poklapaju, sistem za prepoznavanje govora „prepoznaje“ reč i izvršava komandu. Prepoznavanje u zavisnosti od šablona se koristi za relativno male do srednje velike rečnike (uglavnom do nekoliko hiljada reči). Drugi sistemi za zavisno prepoznavanje na osnovu glasa funkcionišu tako što uparuju foneme, više reči i trifone. Pristup fonema/više reči se obično koristi za sisteme sa većim vokabularom, do deset hiljada reči (Xiong, Barker, & Christensen, 2019). Tipično, sistem za zavisno prepoznavanje na osnovu glasa dobro funkcioniše sa srednjim do velikim rečnicima, i sa izolovanim ili povezanim prepoznavanjem reči (Ranjan & Dubey, 2016). Ovi sistemi za zavisno prepoznavanje na osnovu glasa još uvek nisu savršeni i potrebna su poboljšanja, ali nova tehnologija ne bi trebalo da bude ograničena samo na obradu teksta i diktiranje. Postoje mnoge oblasti koje se mogu istražiti i sigurno je da će to mnogi implementirati i usvojiti u godinama koje dolaze.

2. **Prepoznavanje govora koje ne zavisi od obuke sistema** – Sistem sa ovakvom vrstom prepoznavanja govora može da koristi svako, bez potrebe za treningom, odnosno obukom sistema. Mada, tu se susrećemo sa manjim opsegom reči u rečniku, kao i sa većom stopom grešaka. Za računarske sisteme za prepoznavanje glasa nezavisno od obuke, fonemi se izdvajaju iz audio zapisa, pretvaraju u ASCII znakove, a zatim se formulišu u reči kako bi se omogućilo aplikacijama koje koriste prepoznavanje govora da deluju na ulaz (Lattore, Iwano, & Furui, 2005). Postoje matematičke formule i modeli koji se koriste za

identifikaciju najverovatnije izgovorene reči (Dash, Wisler, Ferrari, & Wang, 2019) (Niwatkar & Kanse, 2020). Ovi modeli povezuju izgovorene reči sa poznatim modelima reči i biraju onaj za koji postoji najveća verovatnoća da će biti tačna reč. Da bi se identifikovala „najveća verovatnoća“, velike količine podataka o obuci se koriste za kreiranje modela. Ovaj tip statističkog modela je poznat kao skriveni model Markova, engl. *Hidden Markov model*, HMM) (Keo, Kak, Shiga, Kato, & Kawai, 2019). HMM karakteriše Markov model konačnog stanja i skup izlaznih distribucija. Suština prepoznavanja govora je obuhvaćena u dve vrste varijabilnosti: vremenske varijabilnosti i spektralne varijabilnosti (Miguel, i drugi, 2008). Parametri prelaza su prvi model u Markovljevom lancu, dok drugi model predstavljaju parametri u izlaznoj distribuciji (Fine, Singer, & Tishby, 1998). HMM se zasnivaju na dobrom verovatnosnom uzorkovanju[1] i ima integrisani okvir za istovremeno rešavanje problema segmentacije i klasifikacije (Bargi, Da Xu, & Piccardi, 2017) (poteškoće za računarom u razlikovanju govora od tišine, kako bi segmentirao govor u reči), što ih čini pogodnim za kontinuirano prepoznavanje govornika. U drugim sistemima, gde je detekcija srednje tišine (pauza nekog nerazumljivog izgovora usred govora) otežana, od korisnika se traži da izgovori svaku reč posebno i sačeka da sistem prepozna, što otežava korisnicima da imaju „prirodni“ interfejs za mašinu, interfejs gde se tok razgovora ne prekida prinudnim pauziranjem (Rafieee & Khazaei, 2010). Smatrajući govor uređenom zbirkom fonema, postalo je lako prepoznati govor nezavisno od govornikovog akcenta. Obuka je neophodna, ali u nezavisnim sistemima za prepoznavanje govora, to se radi tako što se model konstruiše korišćenjem velikih uzoraka. Pored HMM-a, koristi se još jedna tehnika usklađivanja uzoraka poznata kao dinamičko vremensko savijanje (Yu, Mason, & Oglesby, 1995). Ova metodologija poredi unapred obrađeni govor sa referentnim šablonom sumiranjem razlika između okvira govora. Neke od reči nisu u skladu sa datim šablonom i tako se neusklađenost ispravlja rastezanjem i kompresijom. Novija tehnika nezavisnog prepoznavanja govora je korišćenje neuronskih mreža (McLaren, Lei, & Ferrer, 2015). Kao što je gore navedeno, HMM tehnologija funkcioniše tako što pravi određene pretpostavke o strukturi prepoznavanja govora, a zatim procenjuje sistemske parametre kao da su strukture ispravne. Ova tehnika može propasti ako su pretpostavke netačne. Neuronske

[1] Verovatnosno uzorkovanje je vrsta uzorkovanja kod kojeg svaki element ciljne populacije ima poznatu, određenu verovatnoću da bude izabran u uzorak. U stvari, količinsko ocenjivanje greške uzorkovanja. https://matematika.pmf.uns.ac.rs/wp-content/uploads/zavrsni-radovi/primenjena_matematika/DunjaArsic.pdf

mreže ne zahtevaju takve pretpostavke. Ovaj pristup koristi distribuiranu reprezentaciju jednostavnih čvorova, čije su veze obučene da prepoznaju govor (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). Za razliku od HMM-a gde znanje ili ograničenja nisu kodirani u pojedinačnim jedinicama, pravilima ili procedurama, već se distribuiraju na mnoge jednostavne računarske jedinice. Neizvesnost se modelira ne kao neverovatnost jedne jedinice, već obrascem aktivnosti u mnogim jedinicama. Ove računarske jedinice su jednostavne prirode, a znanje nije programirano u funkciju bilo koje pojedinačne jedinice; nego leži u vezama i interakcijama između povezanih elemenata obrade.

3. **Sistemi sa diskretnim signalom na ulazu** – Ovakav sistema podrazumeva da osoba koja govori pravi male pauze (oko $^1/_{10}$ sekunde) između reči. Na taj način, omogućava se lakše prepoznavanje gde reč počinje, a gde završava. Ovde se zahteva od korisnika da govori jasno i sporo i da razdvaja reči, što je veoma neprirodan tip govora. Bilo je izveštaja o problemima sa glasom povezanim sa upotrebom diskretnih sistema za prepoznavanje govora (Kambeyanda, Singer, & Cronk, 1997). To je zbog naglog pokretanja i zaustavljanja govora koji je potreban za ove sisteme, zajedno sa monotonim kvalitetom potrebnim za dobro prepoznavanje, a oba su neprirodni obrasci govora.
4. **Sistemi sa kontinuiranim signalom na ulazu** – Osoba može da govori u kontinuitetu, ali softver za prepoznavanje glasa može da prepozna samo određenu količinu reči i fraza. Može se još nazivati „*Word-spotting"* sistem, jer prepoznaje ključne reči dok se govori. I dalje se rade istraživanja ovakvog sistema (Giotis, 2017). Sistem koji koristi tehnologiju kontinualnog signala na ulaza omogućava korisniku da govori na normalniji način bez većih pauza. Brzina unosa je u opsegu normalne brzine ljudskog govora (150 do 250 reči u minuti) (Kong, Somarowthu, & Ding, 2015). Iako je ovim sistemima smanjena mogućnost oštećenja glasnica, ona nije potpuno eliminisana. Kao što je prethodno pomenuto, diskretni sistemi su tačniji za prepoznavanje jedne reči, ponekad se koriste za komande i kontrolu u aplikacijama kao što su tabele i baze podataka. Ali, neki proizvođači (npr. Dragon Systems, Nuance, Inc., Burlington, MA, www.nuance.com) pružaju kontinuirani (npr. Naturally Speaking) i diskretni (npr. Dragon Dictate), u paketu. Sistemi kontinuiranog govora su dizajnirani da razumeju prirodniji način govora. Sistemi za kontinualno prepoznavanje govora zahtevaju velike količine memorije i računare velike brzine. Kako cena ove dodatne računarske funkcionalnosti nastavlja da opada, ovi dodatni zahtevi će biti manje važni.

Simpson (Simpson, 2013) daje detaljan opis ASR-a za pristup računaru, uključujući glavne izazove u primeni ove tehnologije. Kontinuirani govor je sofisticiraniji oblik softvera za prepoznavanje glasa. Daje se mogućnost govorniku da priča prirodno kako bi objasnio svoj problem ili zatražio uslugu. Ovaj program je dizajniran da odabere ključne reči ili fraze i napravi statističku najbolju pretpostavku o tome šta klijent želi. Jasno govorenje pomaže programu da identifikuje potrebe.

5. **Sistemi sa obradom prirodnog jezika** – Ovo je najpoželjniji oblik prepoznavanje govora ili identifikacije na osnovu glasa i svakim danom se sve više razvija. Ovde je korisnik u stanju da slobodno govori, a sistem je u stanju da tumači govor i da izvršava komande u toku izgovora. Sve više je istraživanja u domenu primene Internet stvari gde se vrši međuumrežavanje fizičkih objekata (kuća, automobil,…). Novija istraživanja se svode na mogućnost korišćenja sistema Internet stvari koji može da razume glasovne komande korišćenjem prirodnog jezika (Alexakis, 2019). Korišćenjem prirodnog jezika, kućni uređaji su jednostavniji za upotrebu, te ih je lakše kontrolisati. Čak i kada se da se naredba ili pitanje razlikuje od prethodnih, sistem je u stanju da razume želje korisnika i da odgovori na njih. Ovakav sistem prepoznavanja govora i govornika identifikuje i tumači reči i fraze u govornom jeziku i pretvara ih u tekstove pomoću računara. Obrada prirodnog jezika jednostavno se bavi interakcijom između ljudi i računara koristeći prirodni jezik kao što je npr. u ovom radu srpski i engleski. Tehnologija obrade prirodnog jezika primenjuje algoritme mašinskog učenja na tekst i govor (Kamath, Liu, & Whitaker, 2019). Obrada prirodnog jezika (NLP) omogućava mašinama da razgrađuju i tumače ljudski jezik. To je srž alata koje koristimo svaki dan – od softvera za prevođenje, čatbotova, filtera za neželjenu poštu i pretraživača, do softvera za ispravljanje gramatike, glasovnih asistenata i alata za praćenje društvenih medija. Danas kompanije koriste različite tehnike NLP-a kako bi analizirali objave na društvenim mrežama i znaju šta kupci misle o njihovim proizvodima. Kompanije takođe koriste praćenje društvenih medija da bi razumele probleme sa kojima se njihovi kupci suočavaju korišćenjem njihovih proizvoda.

3. PREGLED PRISTUPA PROBLEMU PREPOZNAVANJA GOVORNIKA

Vrlo često se termin „prepoznavanje glasa“ poistovećuje sa terminom „prepoznavanje govora“. Razlika postoji, jer prepoznavanje govora na osnovu glasa služi za otkrivanje ŠTA je to osoba izgovorila, dok se prepoznavanje glasa bavi odgovorom na pitanje KO je izgovorio određene reči ili rečenice.

Neretko se, u stranoj literaturi, za prepoznavanje glasa koristi i termin „*speaker recognition*“, pa bismo mogli da koristimo i bukvalni prevod „prepoznavanje govornika“. Prema nekim izvorima, prepoznavanje glasa predstavlja svojevrsnu kombinaciju prepoznavanja govora i prepoznavanja govornika (Beigi, 2011).

Ovde je svakako potrebno spomenuti i standarde za prepoznavanje govornika. U oblasti prepoznavanja govornika prisutni su određeni standardi koji su postavljeni od strane različitih internacionalnih organizacija. Ovi standardi imaju veoma važnu ulogu u razvoju i održivosti tehnologije prepoznavanja govornika. Organizacija koja prednjači u definisanju standarda je Američki nacionalni institut za standarde i tehnologije (NIST).

NIST je pokrenuo projekat za unapređenje sistema za prepoznavanje govornika pod nazivom „Procena prepoznavanja govornika“ (eng. *Speaker Recognition Evaluation*)[2]. Cilj ovog projekta je doprinos istraživanjima u oblasti prepoznavanja govornika, preciznije, pronalaženje najboljih algoritamskih pristupa. Projekat je započet još 1996. godine. Procene stanja u ovoj oblasti se vrše na godišnjem nivou posredstvom različitih radionica i konferencija. Tokom godina razvijene su zvanične baze podataka od strane NIST-a koje predstavljaju deo standarda za procenu tekstualno nezavisnih sistema za prepoznavanje govornika. Ciljevi serije evaluacija su (Sadjadi, Greenberg, Singer, Mason, & Reynolds, 2021):

(1) da efikasno izmeri sistem - kalibrisane performanse trenutnog stanja tehnologije,

[2] https://www.nist.gov/itl/iad/mig/speaker-recognition

(2) da obezbedi zajednički okvir koji omogućava istraživačkoj zajednici da istražuje obećavajuće nove ideje u prepoznavanju govornika, i

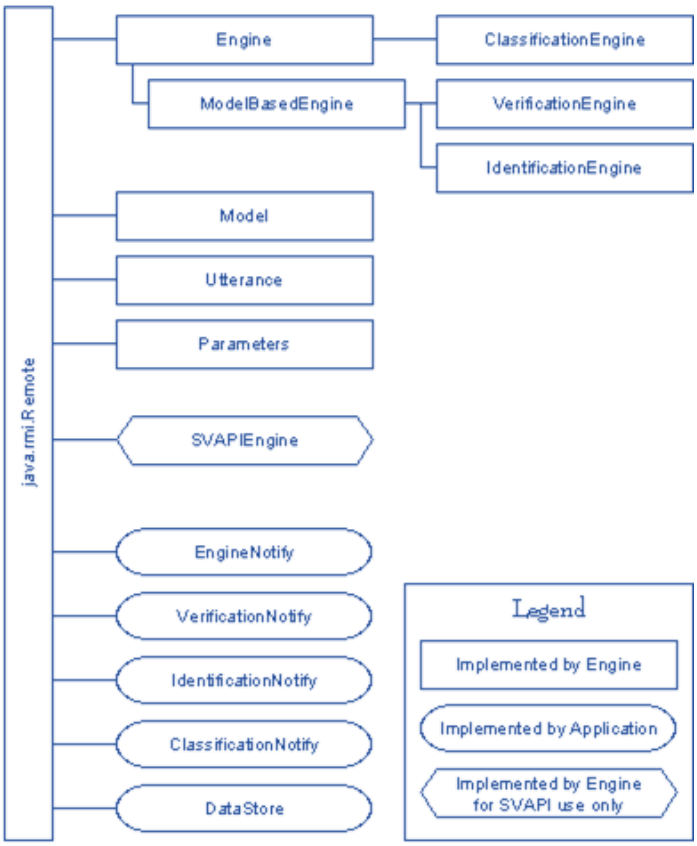
(3) da podrži zajednicu u razvoju naprednih tehnologija koja uključuje ove ideje.

Ideja o istraživanju na tri uređaja, na dva jezika i na četiri programska rešenja otvorenog koda se javila nakon pregleda rada o prikazu procenta neusklađenosti uređaja pri prepoznavanju govornika (Ma, Meng, & Mak, 2007).

Veoma bitnu ulogu u standardizaciji tehnologije prepoznavanja govornika zauzima i API pod naslovom *Speaker Verification Application Program Interface* (SVAPI) (Markowitz, 1999), koji koriste kreatori aplikacija sa tehnologijom za prepoznavanje govornika i na taj način postižu kompatibilnost i interoperabilnost između različitih distributera i proizvođača. Prva verzija ovog standarda je objavljena 1997. godine i ona je predstavljala prvi API standard u oblasti biometrije.

SVAPI obezbeđuje standardne interfejse i klase za aplikacije i specifične programske sisteme, namenska programska jezgra ili engl. *engine*, kako bi mogli da međusobno komuniciraju. I aplikacije i *engine*-i su odgovorni za implementaciju SVAPI interfejsa. Ovi interfejsi definišu metode koje aplikacije i *engine*-i koriste za međusobno komuniciranje, kao i izuzetke koji se mogu dogoditi. U SVAPI specifikaciji definisani su interfejsi i klase u 3 grupe:

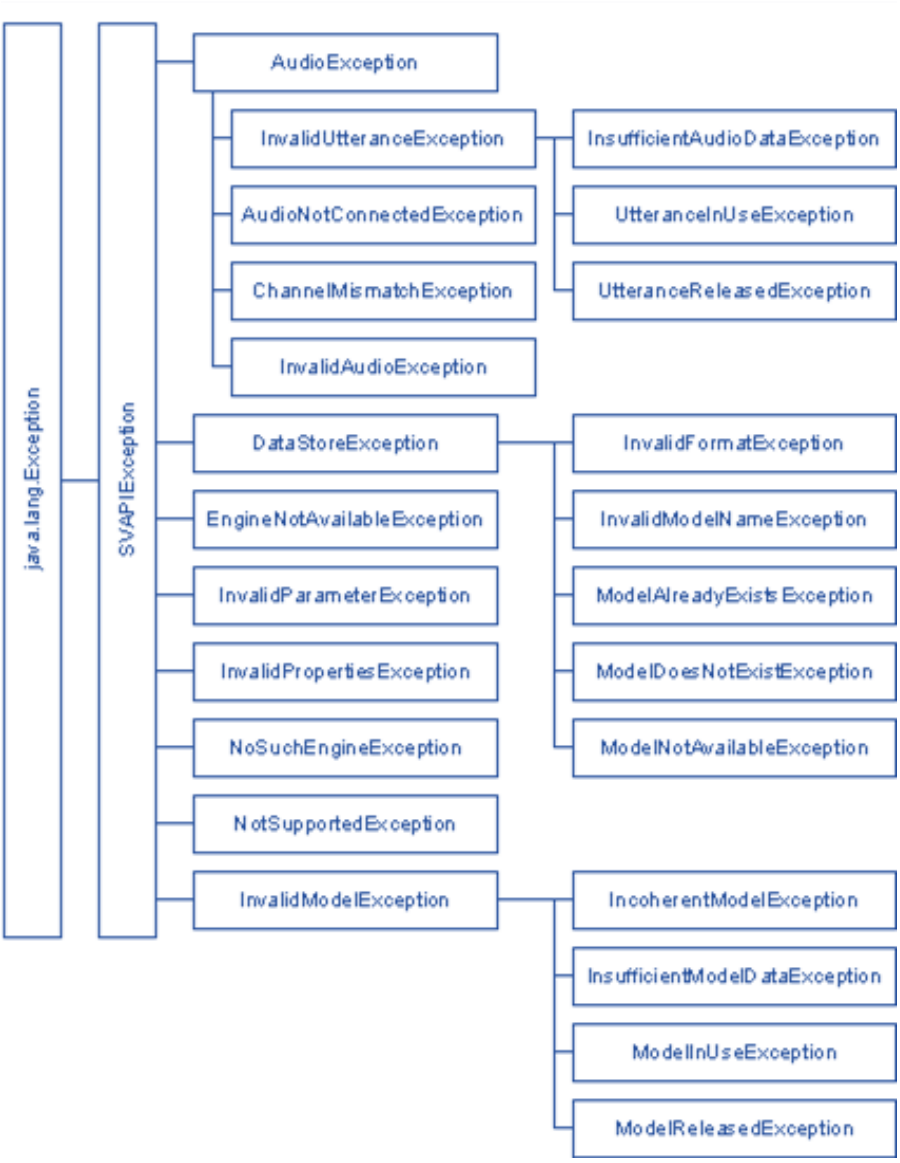
1. Konkretne klase,
2. Interfejsi za namenska programska jezgra,
3. Interfejsi za aplikacije.



Slika 3 Dijagram interfejsa dat u SVAPI specifikaciji

Takođe, SVAPI sadrži i veliki broj klasa koje predstavljaju izuzetke. Na slici 3 prikazan je dijagram ovih klasa kako je dat u SVAPI specifikaciji[3] (Markowitz J. , 2008).

[3] Detaljna specifikacija kao i više informacija o SVAPI standard može se naći na stranici http://www.jmarkowitz.com/downloads.html



Slika 4 Dijagram izuzetaka dat u SVAPI specifikaciji

Prethodno je detaljno objašnjen standard SVAPI, a na slici 4 su prikazani neočekivani ili neželjeni događaji koji se mogu desiti u vreme kompajliranja koda SVAPI aplikacije (Markowitz J. , 2008). Ovo su izuzeci, odnosno greške (eng. *exception*) koje remete normalan tok instrukcije programa.

3.1. TEKSTUALNO ZAVISNI I TEKSTUALNO NEZAVISNI SISTEMI ZA PREPOZNAVANJE GOVORNIKA

Sistem prepoznavanja govornika koristi govorni iskaz kako bi se utvrdilo da li je izrečen od strane poznate osobe. Ovo predstavlja težak zadatak, jer zavisi od više faktora, kao što su kvalitet uređaja

za snimanje, uslov kao i saradnje subjekta. Generalno, prvi zadatak je da se iz snimljenog materijala izvuku relevantne informacije (govorni frejmovi) i uklone manje bitne ili nebitne informacije (tišina, buka, muzika ili govor u pozadini), pre nego što se aktivira algoritam za izdvajanje karakteristika. Mogu se koristiti različiti scenariji za prepoznavanje govornika, zasnovani na tekstualno zavisnim ili tekstualno nezavisnim metodima provere.

Postoje dva glavna pristupa u prepoznavanju govornika (Reynolds D. , 2002):

- Tekstualno zavisno prepoznavanje govornika (eng. *text-dependent speaker recognition*)
- Prepoznavanje govornika nezavisno od teksta (eng. *text-independent speaker recognition*)

Kod tekstualno zavisnog prepoznavanja govornika, tekst koji osoba govori u fazi testiranja mora biti isti kao tekst koji je korišćen u fazi treninga. Tekst može biti isti za sve osobe u bazi sistema ili svaka osoba može imati svoju tekstualnu frazu kao neki vid šifre ili lozinke. U tekstualno zavisnom metodu prepoznavanja govornika, bitan je leksički sadržaj izgovorene reči ovaj metod omogućava bolju robusnost protiv ponovnih napada (Tur & De Mori, 2011). Međutim, ovakvim sistemima je potrebno više resursa nego onima koji se vrše na osnovu tekstualno nezavisnog metoda, kako bi se informacije mogle efikasno leksički procesuirati.

S druge strane, u tekstualno nezavisnom prepoznavanju govornika ne postoje ograničenja koja se tiču reči i rečenica koje osobe mogu koristiti. Ovakvi sistemi se isključivo oslanjaju na fiziološke i bihejviorističke karakteristike glasa određene osobe. Postoji veliki broj softvera koji implementiraju algoritme za prepoznavanje govornika. Mnogi od ovih softvera su vlasnički, međutim sve je više softverskih rešenja za prepoznavanje govornika koji su otvorenog koda. U nastavku rada biće prikazana najpopularnija rešenja otvorenog koda i njihove karakteristike. U tekstualno nezavisnom metodu prepoznavanja govornika, modeli identiteta ne treba da budu zavisni od precizno izgovorenih reči.

Kod prepoznavanja govornika, važno je spomenuti da se za razliku od drugih biometrijskih pristupa, može utvrditi identitet (identifikacija i verifikacija) na daljinu, putem određene infrastrukture, na primer, putem telefona. Ovo čini ovu disciplinu dostupnom u realnom svetu aplikacija, pogotovo sa napretkom računarskih tehnologija (Sadjadi, i drugi, 2017).

Cilj prepoznavanja govornika je da mašina bude u stanju da „čuje“, „razume“ i „postupi“ po verbalnim informacijama. Klevensand i Rodman (R.Klevansand, Voice Recognition, 1997) su spomenuli Davida i ostale istraživače kako su u ranim pedesetim godinama u „*Bell*“ laboratorijama otkrivali sisteme za prepoznavanje govornika, na način gde se izdvaja i prepoznaje izolovana reč ili slog. Cilj je bio napraviti automatsko prepoznavanje govornika putem analize, ekstrakcije podataka, a na kraju i identifikacija govornika. Podaci o govorniku sadrže različite vrste informacija koje identifikuju samog govornika. Tu su uključene i konkretne informacije o govorniku, kao što su govorni trakt i bihejvioralne karakteristike. Analiza govora je u korespodenciji sa segmentima govornih signala za analizu i ektrakciju podataka (Gin-der Wu, 2008).

Metode tekstualno zavisnog prepoznavanja govornika i tekstualno nezavisnog prepoznavanja govornika imaju problem sa šumom u pozadini glasovnog zapisa, jer mikrofon beleži i registruje neželjeni signal kao potencijalnog govornika. Kako bi se ovaj problem rešio, postoje metode koje smanjuju opseg reči i prepoznaju ključne reči (Sadoki Furuki, 2005). Takođe, bitno je spomenuti i metod potpornih vektora koji je postao popularan i moćan alat u tekstualno nezavisnom prepoznavanju govonika, jer u jezgru ovakvog metoda se daje izbor ekspanzije karakteristika (Asghar Taheri, February 2005).

El-Moneim sa saradnicima (El-Moneim, i drugi, 2021) je predstavio dva nova pristupa za tekstualno zavisno prepoznavanje govornika, kao i za tekstualno nezavisno prepoznavanje govornika. U sistemu za tekstualno zavisno prepoznavanje govornika predložili su robustniji pristup od klasičnog testa odnosa verovatnoće (engl. *likelihood ratio test*, LRT). Njihov algoritam koristi hibridni generativno-diskriminativni okvir u kojem će generativni model naučiti karakteristike govornika, a zatim će diskriminativni model da napravi razliku između govornika i varalice. Jedna od prednosti predloženog algoritma je u tome što ne zahteva da se ponovo obuči generativni model. Predloženi model, u proseku, daje 36,41% relativno poboljšanje vrednosti parametra EER (*eng. Equal Error Rate*) u odnosu na LRT. Što se tiče sistema za tekstualno nezavisno prepoznavanje govornika, urađeno je novo tumačenje Univerzalnog pozadinskog modela (engl. *Universal background model*, UBM) i smatraju ga funkcijom mapiranja koja transformiše posmatranja promenljive dužine (govorne izjave) u vektor obeležja fiksne dimenzije (dovoljna statistika). Nakon ovog mapiranja, merenje sličnosti se izračunava na karakteristikama

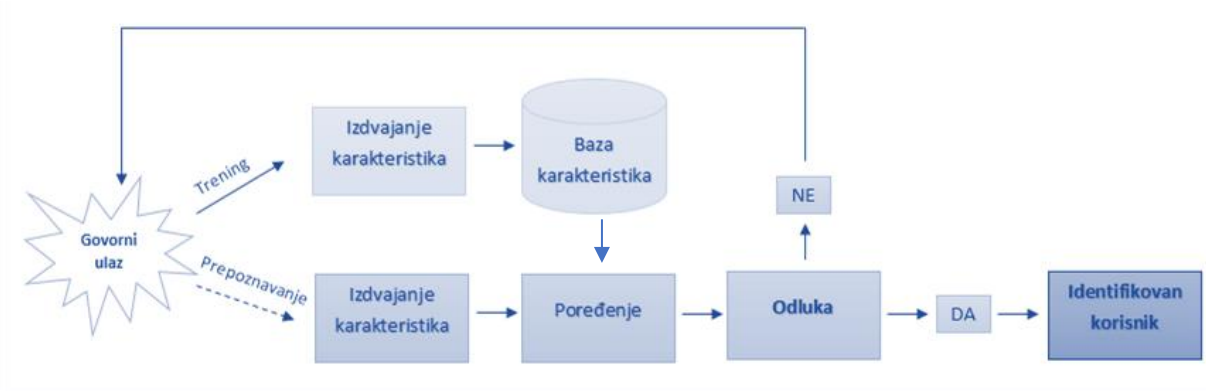
fiksnih dimenzija. Sa ovom novom interpretacijom, predložili su novo merenje sličnosti koje proizvodi više od 10% relativnog poboljšanja u odnosu na konvencionalni UBM-MAP, (engl. *maximum a posteriori*, MAP) okvir u jednakoj stopi greške i funkciji troškova detekcije.

3.2. OPŠTA ARHITEKTURA SISTEMA ZA AUTOMATSKO PREPOZNAVANJE GOVORNIKA

U procesu prepoznavanja govornika se spominju dve oblasti, identifikacija i verifikacija. Identifikacijom se utvrđuje identitet osobe tokom izgovora reči/rečenica. Proces verifikacije predstavlja proveru identiteta.

Prepoznavanje govornika se odnosi na prepoznavanje osoba na osnovu njihovog glasa. Svaka osoba ima sebi svojstven glas i to zbog različitih fizičkih, ali i bihejvioralnih karakteristika, kao što su akcenat, ritam, način izgovaranja...

Na slici 5 je predstavljena opšta arhitektura sistema za prepoznavanje govornika u režimu upisa/obuke/treninga i u režimu ispitivanja/verifikacije/prepoznavanja. U režimu treninga novi govornik (sa poznatim identitetom) se upisuje u bazu podataka, dok u režimu prepoznavanja nepoznati govornik snima govorni ulazni signal i sistem pokušava da identifikuje govornika. Ovaj sistem se može koristiti i za identifikaciju i za verifikaciju.



Slika 5 Opšta arhitektura sistema za prepoznavanje govornika

Neka istraživanja (Saquib, Salam, Nair, Pandey, & Joshi, 2010) navode kako postoje dve glavne komercijalizovane primene tehnologija i metodologija za prepoznavanje govornika: identifikacija govornika i verifikacija govornika (objašnjeno već u poglavlju 2.2.). Takođe, oni spominju da

postoje različite arhitekture sistema za prepoznavanje govornika koje obezbeđuju istraživači i programeri.

3.3. IZVOĐENJE KARAKTERISTIKA GLASA

Glas predstavlja individualno obeležje svakog čoveka i njegova osnovna funkcija je komunikacija sa drugim ljudima. On može da služi za dokazivanje identiteta subjekta proverom, na primer, izgovora unapred određene fraze. Spada u grupu nenametljivih tehnika. Međutim, glas je artefakt podložan promenama. Od ove karakteristike ne očekuje se da bude jedinstvena. Verodostojna i neupitna identifikacija govornika treba biti zasnovana na egzaktnim, naučno utemeljenim parametrima. Kada se govori o identifikaciji osobe putem glasa, onda gledamo karakteristike poput boje glasa, modulacije, frekvencije, specifičnosti izvora određenih glasova, govornim manama i drugo. Ovaj način identifikacije pripada području fonoskopske identifikacije.

Eksperti za ovu oblast posebnim uređajem glas ispitanika pretvaraju u električne signale, karakteristike vizualizuju posredstvom grafikona, pa potom vrši upoređivanje sa referentnim govornikom. Shvatanje eksperata o stepenu podudarnosti ili nepodudarnosti glasa može se smestiti u pet kategorija (Cai, Qin, & Li, Multi-Channel Training for End-to-End Speaker Recognition Under Reverberant and Noisy Environment, 2019):

1. Definitivno je u pitanju isti govornik,
2. Definitivno nije isti govornik,
3. Verovatno isti govornik,
4. Verovatno nije isti govornik,
5. Otežana identifikacija radi lošeg snimka.

Savremena tehnologija omogućava znatno podizanje nivoa kvaliteta snimka, odnosno restituciju loših snimaka i drugo (Nimac, 2007).

Pri upoređivanju raznih dostupnih biometrijskih metodea veoma važno je imati valjane kriterijume. Set kriterijuma vrednovanja, koji razvili su eksperti za biometriju, sadrži (Jain A. K., 1999):

- **Univerzalnost** – odnosi se na udeo osoba koje poseduju osobinu neophodnu za autentifikaciju. U savršenom scenariju taj udeo je 100%.

- **Jedinstvenost** – označava potrebu da svaka osoba u skupu ima jedinstvenu registrovanu biometrijsku karakteristiku. Metode otiska prsta, dužice oka ili termogram lica imaju veliku sposobnost razlikovanja karakteritika, tako da verovatnoća da dve osobe imaju jednake regostrovane karakteristike je bliska nuli.
- **Trajnost** – posmatrana karakteristika ne bi smela da se menja sa vremenom. Neke karakteristike, kao dužica oka ili termogram lica, tokom vremena ostaju nepromenjene, ali neke, kao slika lica, glas ili karakteristike ponašanja starenjem se menjaju.
- **Prikupljivost** – odnosi se na pitanje da li se karakteristika može lako izmeriti i kvantitativno izraziti. Primeri takvih metoda su potpis, kao i prepoznavanje i termogram lica. Jedna od slabih prikupljivih metoda je skeniranje mrežnjače.
- **Efikasnost** – označava tačnost i brzinu biometrijske metode. Tačnost se može kvantitativno prikazati merama, kao što su FAR, FRR i EER. Veliku pažnju treba obratiti i na brzinu kojom se obavlja upoređivanje. Zbog velikog broja upoređivanja koje se provode, sporije metode nisu pogodne za identifikaciju. Tačnost i brzina mogu se izraziti kao posebna svojstva.
- **Prihvatljivost** – kriterij koji označava spremnost korisnika da se podvrgnu postupku prikupljanja date biometrijske karakteristike. Sistemi koji skeniraju mrežnjaču, zahtevaju prodor laserskog snopa u dubinu oka korisnika, što je nametljivo i ne doprinosi popularnosti metode.
- **Mogućnost zaobilaženja -** otkriva nam kako jednostavno se može prevariti sistem korišćenjem određenih tehnika.
- **Skalabilnost** – odnosi se na fleksibilnost sistema, tj, na mogućnost usklađivanja sistema zahtevima tokom vremena korišćenja, a da tom prilikom se bitno ne ugrozi efikasnost sistema.
- **Zrelost** biometrijske metode označava dovoljnu praktičnost metode za primenu u realnom okruženju.
- **Troškovi** biometrijskih sistema su važan faktor. Postoji podela na inicijalne troškove opremu (hardver, softver), kao i na troškove eksploatacije, licenciranja i održavanja.

Za najbolje i najpreciznije rezultate rada sistema za prepoznavanje glasa, sistem treba da zadovoljava sledeće kriterijume (Wolf, 1972):

- ima manju varijabilnost u okviru govornika, a veću između različitih govornika,
- stabilan tokom vremena,
- prepoznaje imitaciju glasa,
- robustan na probleme u prenosu i šum,
- jednostavno izdvajanje i merenje i veliki broj obrada uzoraka glasa.

Karakteristike glasa se mogu, u pogledu informacija koje sa sobom nose, podeliti u tri grupe:

1. Karakteristike visokog nivoa (fonemi, akcenat, semantika, izgovor)
2. Spektralno-temporalne karakteristike (intonacija, energija, ritam, trajanje, vremenske karakteristike)
3. Kratkoročne spektralne karakteristike (spektar)

Od posebnog interesa za ocenu kvaliteta rada biometrijskog sistema su sledeće metrike za merenje performansi (Monroe, 2012):

- *false accept rate* ili *false match rate* (FAR ili FMR) – verovatnoća da sistem na osnovu ispitivanog uzorka i baze registrovanih biometrijskih uzoraka pogrešno zaključi da postoji potrebno preklapanje sa nekim uzorkom u bazi.
- *false reject rate* ili *false non-match rate* (FRR ili FNMR) – verovatnoća da sistem na osnovu ispitivanog uzorka i baze registrovanih biometrijskih uzoraka pogrešno zaključi da ne postoji potrebno preklapanje sa nekim uzorkom u bazi.
- *receiver operating characteristic* ili *relative operating characteristic* (ROC) – predstavlja meru kvaliteta rada biometrijskog sistema posmatran mogućim odnosima između FAR i FRR, kao skup mogućih radnih tačaka sistema,
- *equal error rate* ili *crossover error rate* (EER ili CER) – karakteristična radna tačka sistema u kojoj su i iste stope prihvatanja greške i odbijanja greške rada sistema. Što je vrednost ERR parametra manja, to je sistem pouzdaniji u donošenju zaključaka.
- *failure to enroll rate* (FTE ili FER) – verovatnoća nemogućnosti registrovanja uzorka iz prikupljenog materijala.

- *failure to capture rate* (FTC) – verovatnoća nemogućnosti uzimanja sirovih biometrijskih podataka ispitanika
- *template capacity* – maksimalni broj zapisa, registrovanih šablona, u bazi biometrijskih podataka.

3.4. MODELI GOVORNIKA

Osnovni cilj modela govornika je da sistem bude u stanju da na osnovu govora nekog govornika odredi jedinstven skup vrednosti parametara modela govornika, kao identifikatora koji ga dovoljno razlikuje od svih drugih govornika u bazi biometrijskih podataka.

Još 1995. godine predstavljeni su osnovi modela na kojima se i danas baziraju drugi nadograđeni modeli (Reynolds D. , 1995):

- Statistički model za prepoznavanje govornika - Govornik je modeliran kao nasumični izvor koji proizvodi posmatrane vektore karakteristika. Unutar slučajnog izvora nalaze se stanja koja odgovaraju karakterističnim oblicima vokalnog trakta.
- Model govornika baziran na Gausovim modelima mešavine, engl. *Gaussian Mixture Models*, GMM - Ovi modeli dosta preciznije opisuju i porede intra-spikerske karakteristike, prilikom upoređivanja uzorka, od metoda baziranih na vrednostima visine tona (engl. *Pitch*) i formantima.
- Skriveni Markovljevi modeli, HMM - Ovi modeli imaju najveću primenu pri prepoznavanju govornika na osnovu sadržine govora.
- Sada su popularni modeli obrade prirodnog jezika (NLP) obučeni na podacima iz međuljudske komunikacije. Oni mogu biti nepouzdani jer, bez ikakvih ograničenja, mogu da uče iz lažnih korelacija koje nisu relevantne za postavljeni zadatak. Pretpostavljamo da obogaćivanje modela informacijama o govornicima na kontrolisan način može da vodi do relevantnih induktivnih zaključaka. Istražuje se koje su to fraze koje mogu uticati na prepoznavanje uzorka, te se razvijaju personalizovani modeli (Ostapenko, Wintner, Fricke, & Tsvetkov, 2022).

Algoritmi koji se koriste za verifikaciju glasa govornika mogu se podeliti u sledeće kategorije (Reid, 2004):

- *verifikacija s fiksnom frazom* (eng. *Fixed phrase verification)* - Kod ovog algoritma, koristi se samo jedna reč za upis i proveru. Metod provere poklapanja je jednostavan, jer se porede dve talasne forme. Ukoliko su dve forme u dozvoljenom nivou tolerancije, smatra se da je došlo do poklapanja podataka.
- *verifikacija s fiksnim rečnikom* (eng. *Fixed vocabulary verification*) - Ovaj algoritam koristi skup reči koji se sastoji od cifara (0-9) i nasumično povezanih reči. Prilikom upisa, korisnik izgovara svaku od reči u rečniku kako bi se kreirao jedinstven korisnički model. Prilikom verifikacije, korisnik treba da izgovori neku od reči iz rečnika. Tada se ta reč proverava u modelu za tog korisnika, te se vrši poređenje.
- *verifikacija s fleksibilnim rečnikom* (eng. *Flexible vocabulary verification*) - U ovom slučaju, prilikom faze upisa, korisnik treba da ponovi proizvoljan broj reči iz ponuđenog leksikona. Preduslov je da se svaki fonem[4] leksikona nalazi bar u jednoj odabranoj reči. Takođe, međusobni odnos fonema mora biti adekvatan. Prilikom verifikacije, korisnik izgovara bilo koju od reči iz leksikona. Ta reč se razbija na foneme i onda se vrši poređenje.
- *verifikacija s tekstualno nezavisnim metodom* (eng. *Text-independent verification*) - Kod ovog algoritma korisnik ima slobodu da izabere bilo koju frazu ili reč za potvrdu identiteta.

Postoje tri modela strukture govora, koja se koriste u svrhe prepoznavanja:

- *Akustični model* sadrži akustična svojstva svakog senona. Postoje i modeli nezavisni od konteksta koji sadrže svojstva (najverovatnije karakteristične vektore za svaki glas) i modeli zavisni od konteksta (izgrađeni od senona sa kontekstom).

- *Fonetski rečnik* sadrži mapiranje iz reči u glasove. Ovo mapiranje nije posebno efikasno, jer se beleže samo dva do tri načina izgovaranja, ali fonetski rečnik je dovoljan i u većini vremena služi svrsi. Proces mapiranja reči u glasove, takođe se može izvesti i preko kompleksnih funkcija koju bi mašina mogla da nauči preko algoritama učenja.

[4] **Fonem** je najmanja jedinica govora koja je bitna za značenje.

- *Jezički model* koji se koristi da bi se ograničila pretraga mnoštva reči. Ovaj model u kontekstu prethodnih definiše koja reč bi mogla da bude sledeća, te na taj način ubrzava pretraživanje isključujući reči koje nisu verovatne. Da bi se postigla visoka preciznost, jezički model mora uspešno da eliminiše skup reči koje nisu verovatne. U stvari, on mora biti dobar u predviđanju sledeće reči u sekvenci.

3.5. PRISTUPI PROBLEMU ROBUSTNOSTI

Robustnost u prepoznavanju govora se odnosi na potrebu da se održi dobra preciznost prepoznavanja, čak i kada se degradira kvalitet ulaznog govora, kao i kada se izvedene govorne karakteristike razlikuju u treningu i operativnom korišćenju. Robustnost mnogo zavisi od postavke sistema za prepoznavanje. Prepoznavanje govora je, takođe, veoma podložno greškama i u slučajevima pogrešnog izgovora ili izražavanja, jer glas predstavlja bihejvioralnu biometriju.

Osnovni razlog za važnost razmatranja problema robustnosti u programskim sistemima koji koriste biometrijske podatke potiče od prirode ulaznih podatka, kao predmeta rada biometrijskih sistema. Za razliku od konvencionalnih programskih sistema u kojima su ulazni podaci potpuno struktuirani i jednoznačno interpretabilni, biometrijski sistemi pripadaju klasi multimedijalnih informacionih sistema u kojima su ulazni podaci „sirovi“, semistruktuirani, odnosno nepotpuno interpretirani, u formi zvučnog zapisa, slike ili video zapisa (Starčević & Štavljanin, 2013)[5]

Pošto se biometrijski podaci u procesu akvizicije preuzimaju od stvarnih, fizičkih osoba, oni prolaze kroz nekoliko faza obrade.

U prvoj fazi se koristi analogno-digitalni konvertor (ADC), sa kojim se meri izvesna fizička ili fiziološka osobina čoveka i pretvara u računarski čitljiv oblik, digitalni podatak. U opštem slučaju, i u idealnom slučaju već se u ovoj fazi se za svaki multimedijalni entitet unose dve sistemske greške: greška kvantizacije (engl. *Quantization error*), koja zavisi od veličine rezolucije konvertora i broja prostornih dimenzija, i greška uzorkovanja (engl. *Sampling error*), odnosno greška vremenskog odabiranja. Ovakav računarski čitljiv oblik podatka se zove „sirov“ podatak. Pored toga, „sirov“ podatak sadrži i neželjene vrednosti signala „smetnje“ ili „šuma“, jer praktično nije moguće da se iz realnog sveta posmatra i izdvaja samo koristan signal, u našem slučaju samo glas

[5] Starčević, D., Štavljanjin, V., *Multimediji,* Fakultet organizacionih nauka, Beograd, 2013. str.11

neke osobe. Signalu korisnikovog glasa se superponiraju zvuci iz okruženja, kao što su glasovi drugih ljudi i buke koju stvaraju razni mehanički, električni i elektronski uređaji, zvuci strujanja vazduha, vode, životinja, a i sam ADC uređaj, zbog svoje nesavršenosti, odnosno kvaliteta izrade, unosi specifična linearna i nelinearna izobličenja u koristan signal. Dodatni izvori grešaka se javljaju u nesavršenosti procedura uzimanja biometrijskih podataka, temperature, vlažnosti, osvetljenja, kao i tokom prenosa signala od mesta zahvatanja signala do mesta njegove dalje obrade.

U drugoj, opcionalnoj, fazi obrade, „sirov" podatak se podvrgava predobradi (engl. *preprocessing*), odnosno postupku odstranjivanja signala smetnji kako bi se dobio što „čistiji" signal za sledeću faze obrade.

Pročišćeni „sirov" podatak i dalje je semistruktuiran biometrijski podatak najčešće u formi matrice podataka određenog reda. Matrica eksplicitno sadrži samo deo semantike koju nosi, na primer, fotografija neke osobe ili zvučni zapis govora. Drugi, značajan deo semantike ostaje skriven u podacima, kao, na primer, odgovori na pitanja „Ko je osoba na slici?" ili „Ko je osoba kojoj čujemo glas?", „Na kom jeziku priča?", „Šta priča?" i dalje skriven. Odgovore na postavljena pitanja mogu se dobiti sa izvesnom verovatnoćom u sledećoj fazi obrade ulaznih biometrijskih podatka pomoću programskih sistema poznatih pod generičkim nazivom *Interpreteri.*[6]

Međutim, rad biometrijskog sistema sa „sirovim", pa čak i poboljšanim „sirovim" podacima je izrazito nepraktičan. Naime, biometrijski sistem koji bi radio samo sa „sirovim" biometrijskim podacima sadržavao bi ogromnu količinu podataka, kako zbog velikog broja obuhvaćenih osoba (na primer, populacija na nivou regije ili države!), tako i zbog velike dužine pojedinačnog „sirovog" medijskog zapisa o osobi (engl. *Binary Large Object - BLOB*). Iako danas spoljni računarski mediji mogu imati terabajte korisničkih podataka, pristup i brzina rada sa takvim podacima bi značajno uticala na nefunkcionalne performanse biometrijskog sistema. Stoga je uobičajeno da operativna baza biometrijskih podatka umesto originalnih „sirovih", sadrži pomoću interpretera izvedene (engl. *Extracted*) biometrijske podatke. Izvedeni biometrijski podaci u dovoljno velikoj meri sadrže specifične vrednosti parametara odabranog modela posmatrane biometrijske karakteristike, na primer, odabranog modela glasa govornika ili crta lica osobe sa

[6] Starčević, D., Štavljanjin, V., *Multimediji,* Fakultet organizacionih nauka, Beograd, 2013. str.11

fotografije. Naravno, da su izvedeni zapisi posmatranog biometrijskog svojstva – *šabloni,* potpuno struktuirani, da zauzimaju nekoliko redova veličine manje potrebnog memorijskog prostora i da tako omogućavaju mnogo brži rad biometrijskih sistema u režimu pretraživanja baze podataka i donošenja odgovarajućih odluka.

Ipak, iako rad sa šablonima biometrijskih karakteristika znatno ubrzava rad biometrijskog sistema, zbog same prirode biometrijskih podatka kao medijskih podatka, sistem je i dalje osetljiv na opisane uslove akvizicije biometrijskih podatka, odnosno načina interakcije korisnika sistema u postupku formiranja baze biometrijskih podataka i u režimu eksploatacije. Istraživanja koja se i dalje vrše, pokušavaju da pronađu modele biometrijskih karakteristika koji su dovoljno rezolutni sa izvedenom vrednošću karakteristike osobe i za slučaj velike populacije, dovoljno jednostavni za brzo izvlačenje potrebnih vrednosti parametara za konkretnu karakteristiku osobe, koje zauzimaju malo fizičkog prostor za smeštaj izvedenih vrednosti, pa omogućavaju brzo procesiranje i pretraživanje baze karakteristika, ali i koje su slabo osetljive na varijacije u načinu interakcije, korišćenim ADC uređajima, osobinama okruženja u kojem se interakcija vrši, jeziku komunikacije, dužini trajanja komunikacije, emocionalnom i zdravstvenom stanju korisnika i prenosnim vezama od mesta akvizicije do mesta obrade (Rao & Sarkar, 2014).

Poseban problem u vezi robustnosti sistema pojavljuje se u slučajevima korišćenja različitih uređaja za unos glasa u sistem, u različitim radnim okruženjima, u programskim sistemima za identifikaciju govornika razvijenim za jedno govorno područje, a koje treba lokalizovati za drugo govorno područje, na primer kod aplikacija e-trgovine (Misra & Hansen, 2018), (Ma & al, 2007).

Togneri i Pulela (Togneri & Pullella, 2011) sumirali su aplikacije za robusnu identifikaciju govornika. Praktični sistemi za prepoznavanje govornika često su podložni šumu ili izobličenjima unutar ulaznog govora što degradira performanse. U sistemima koji se koriste telefonsku komunikaciju u kancelarijskom okruženju, glavni oblik degradacije signala govora je uzrokovan slušalicama i/ili mikrofonom. Međutim, za prepoznavanje govornika u slučaju kada je govornik udaljen i na otvorenom prostoru, izobličenja do kojih dovodi okolina, pozadinski šum, je zabrinjavajuća. Tipična rešenja za obezbeđivanje robusnosti u prepoznavanju uklanjaju šum iz informacija o karakteristikama govornika i uključuju metode kao što su normalizacija srednje vrednosti (Furui, 1981), RASTA tehnika (Hermansky, Morgan, Bayya, & Kohn, 1991), i robusne parametrizacije (Nikias & Mendel, 1993) (Wenndt & Shamsunder, 1997).

Kod novijih istraživanja nailazimo na postavljanje ograničenja u postupku ekstrakcije karakteristika iz trening uzorka i bučnih kopija kako bi se algoritam naučio robusnom ponašanju za slučaj govora u bučnom okruženju tokom operativnog korišćenju (Cai, Cai, & Li, 2020)

3.6. KRITIČKI OSVRT NA POSTOJEĆE PRISTUPE PROBLEMU PREPOZNAVANJA GOVORNIKA

Treba spomenuti da je ovaj rad rađen zbog aktuelnosti biometrijske tehnologije gde svaka od njih koristi različite biometrijske senzore i različite algoritme za uparivanje biometrijskih podataka.

Nekoliko postojećih alata pomažu istraživačima da izgrade i procene svoje sisteme za prepoznavanje osobe na osnovu glasa. Na osnovu rezultata prethodnih istraživanja se vidi kako funkcioniše HTK alat (Morales, 2005) i SPro3 (*speech signal processing toolkit)* za ekstrakciju podataka i prepoznavanje glasa (Larcher, 2013). Kako bi se izvršila interpretacija zvučnog zapisa govornika, potrebno je izvršiti izdvajanje karakteristika. Govorni signal možemo posmatrati po delovima kao kratkotrajan stacionarni signal, odnosno može se pretpostaviti da u kratkom vremenskom periodu od, na primer, deset milisekundi, govor može biti shvaćen kao stacionaran process (Vaseghi S. , 1996). Prepoznavanje govornika sastoji se od detekcije govornih aktivnosti, ekstrakcije vrednosti suštinskih biometrijskih parametara i normalizacije, dok se u pozadini vrši proces modelovanja, upisa, upoređivanja i konačne odluke. Takav proces se vidi na slici 6 (Kinnunen T. a., 2010) gde su prikazane glavne faze sistema za prepoznavanje govornika. Na slici se vidi da prepoznavanje govornika se odnosi na kategoriju digitalne obrade signala gde govor služi kao obeležje po kojem se osoba razlikuje. Prvi korak je da se podaci koji su se prikupili pošalju putem protokola na upis, trening, odnosno evaluaciju.

U korišćenju razvijenog biometrijskog sistema postoje dve faze (Latinović, 2017). Jedna se odnosi fazu upisa (eng. *enrollment),* gde se vrši snimanje glasa konkretnog ispitanika, izdvajaju sekarakteristike, te se zatim formira uzorak ispitanika. U drugoj fazi koja se zove faza rada govornik priča u mikrofon. Na taj način dobijamo novi uzorak, te se vrši upoređivanje sa uzorkom nastalim u fazi upisa, a nalazi se sad u bazi. Upoređivanjem se dolazi do odgovarajućih rezultata koji označavaju verovatnoću da je govornik onaj koji tvrdi da jeste. Ovim putem vrši se formiranje

nove baze koja će se upotrebljavati u fazi testiranja. Nakon toga se identifikuju podaci, obavlja izdvajanje karakteristika i na kraju normalizacija.

Ovo istraživanje je izvedeno zbog toga da bi se došlo do što boljeg metoda koji bi prepoznavao osobu na osnovu glasovnog zapisa. Takođe, ovde ima i nepraktičnosti, ako se radi o manjku potrebnih informacija i zato je uvek bolje ubaciti što više podataka u bazu, da bi tačnost bila što veća. Kod sistema gde su velike baze podataka, verifikacioni sistemi radi veće brzine rada imaju nadmoć u odnosu na identifikacione sisteme. Takođe, možemo reći da vrednosti grešaka i vremena potrebna za pojedine radnje su veća u stvarnim uslovima, a u vezi sa biometrijskim sistemima Možemo još pridoneti ako nezavisno testiramo, te na taj način garantujemo kvalitetno tržište biometrijskom tehnologijom. Velika prednost ove metode je ta što je jeftina i svima vrlo lako dostupna, mada je mnogi smatraju dopunskom biometrijskom metodom.

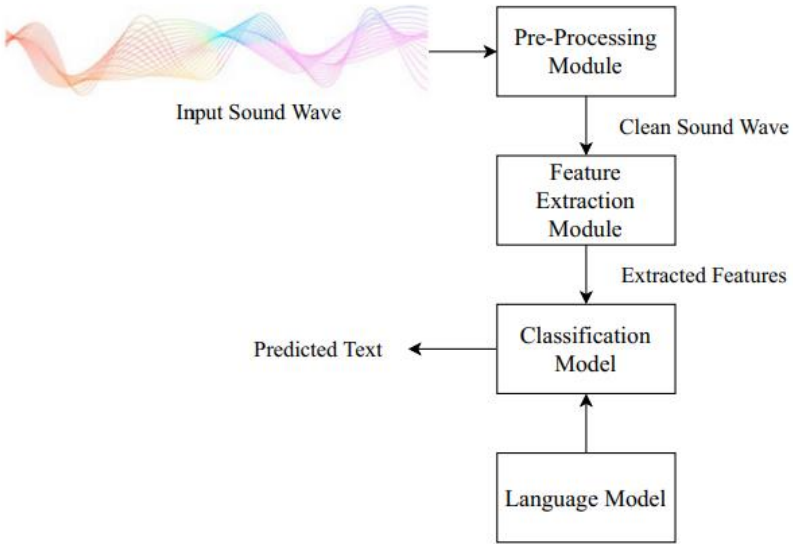
Dat je pregled rešenja otvorenog koda koja se mogu koristiti za implementaciju biometrijskog sistema za prepoznavanje govornika. Ovakva rešenja otvorenog koda su sve više dostupna i već postoji nekoliko softvera koji su se izdvojili svojim performansama i implementiranim algoritmima. Rešenja predstavljena u radu su SPEAR, MARF, ALIZE i HTK, koja su se izdvojila prilikom pregleda literature kao najčešće korišćene platforme otvorenog koda za prepoznavanje osoba na osnovu glasa. Prikazana je specifičnost svakog od navedenih softvera. Takođe je prikazan i model sistema za prepoznavanje govornika zasnovan na rešenjima otvorenog koda.

Najveća prednost korišćenja navedenih tehnologija prilikom implementacije sistema za prepoznavanje govornika je upravo otvoreni kod. Pre svega, ovakav kod je javno dostupan što omogućava velikom broju programera da svakodnevno unapređuju i modifikuju ove softvere. Rešenja otvorenog koda su besplatna što je samo po sebi velika prednost. S obzirom na prisutnost velikog broja ovakvih rešenja za prepoznavanje govornika moguće je realizovati kombinaciju sa više ovakvih rešenja i dobiti kompletniji sistem za prepoznavanje osobe na osnovu glasa sa pouzdanijim donošenjem zaključaka.

Kinnuen i Tomi (Buzo, 2014) su prikazali komponente sistema za automatsko prepoznavanje govornika gde se u gornjoj slici nalazi proces upisa, a u donjoj process prepoznavanja govornika. Procedura izdvajanja korisnih signala iz sirovih signala se naziva model izdvajanja karakteristika. izdvajanje parametara (osobina). U fazi upisa, model govornika je obučen da koristi za ciljanog

govornika. U fazi prepoznavanja govornika, izdvajanje karakteristika (od nepoznate osobe) se upoređuju sličnosti sa podacima u bazi podataka. Konačna odluka se donosi nakon upoređivanja ovih sličnosti.

Osnovna struktura sistema za automatsko prepoznavanje govornika prikazana je na slici 6. Ovakav proces prepoznavanja govornika osmislio je Malik sa kolegama.



Slika 6 Sistem za automatsko prepoznavanje govornika (Malik, Malik, Mehmood, & Makhdoom, 2021)

4. PREGLED PROGRAMSKIH REŠENJA ZA PREPOZNAVANJE GOVORNIKA SA OTVORENIM PROGRAMSKIM KODOM

U daljem istraživanju realizovan je sistem korišćenjem modela predstavljenog u ovom radu i izvršeno je njegovo testiranje. Dalje, ispitane su mogućnosti svakog od predstavljenih softvera i videli smo gde se uklapaju u predstavljenom modelu. Uvidom u navedene činjenice možemo zaključiti da u ovoj oblasti postoji dosta prostora za istraživanje primene softvera otvorenog koda u implementaciji biometrijskih sistema koji koriste prepoznavanje glasa. Za ovo istraživanje odabrani su softveri otvorenog koda zbog toga što su lako dostupni, zastupljeni su u praksi, mogu da koriste različite algoritme, te daju adekvatnu podršku korisniku. Navešćemo još neke prednosti:

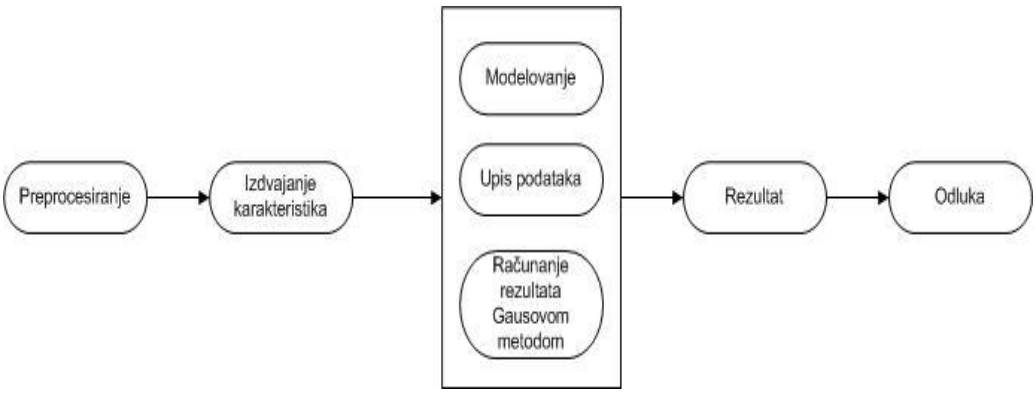
- Pomažu kompanijama da uštede vreme i novac mehanizacijom poslovnih procesa. Npr. tokom telefonskih poziva, pruža trenutne poglede na ono što se dešava.
- Isplativiji su jer softver obavlja zadatak prepoznavanja i transkripcije govora brže i tačnije od čoveka.
- Cena softvera – besplatno. Besplatni softver za prepoznavanje govora dostupan je u mnogim oblicima kao što su veb, mobilni i desktop.
- Jednostavan je za upotrebu i lako dostupan. Npr. u računarima i mobilnim uređajima, softver za prepoznavanje govora se često instalira u računare i mobilne uređaje koji omogućavaju lak pristup.
- Mogu efikasno da ispune naše zahteve. Npr. prepoznavanja govornika u e-aplikacijama.
- Mogućnost pravljenja aplikacije za prepoznavanje govornika koja zahteva napredne tehnike obrade govora.
- U zavisnosti od softvera za prepoznavanje govora otvorenog koda, možete koristiti prepoznavanje govora da biste razgovarali sa računarom, čitali dokumente, otvarali, uređivali i slali e-poštu.

4.1. SPEAR

Spear je open-source rešenje i predstavlja prošireni alat za prepoznavanje govornika. Baziran je na *Bob*-u koji podrazumeva slobodnu obradu signala i mašinskog učenja (A. Anjos, 2012).

SPEAR[7] je alat za prepoznavanje glasa zasnovan na Bob-u. Bob je kreiran da zadovolji potrebe istraživača i zasnovan je na efikasnoj C++ implementaciji velikog seta algoritama za mašinsko učenje i obradu podataka. Bob, takođe, pruža researcher-friendly Pyhton okruženje, koje u odnosu na druge skražuje istraživaču vreme koje ulaže u razvoj (development).
Bob se oslanja na otvorenu i prenosivu (portable) HDF5 biblioteku i formatiranje fajlova za skladištenje i obradu podataka. Ceo kod je dobro dokumentovan i testiran kroz nekoliko platformi.[8]

Spear uzima prednosti stečenog iskustva od facereclib10[9], open source platforme za prepoznavanje lica koja ima za cilj da uporede različite varijacije state-of-the-art algoritama nad nekoliko baza podataka sa slikama lica. Slično kao facereclib, Spear uključuje set podesivih command line skripti, koje dozvoljavaju pokretanje na više različitih platformi (GMM, JFA, ISV i i-vectors zasnovanim sistemima) na well-known ili in-house bazama za prepoznavanje glasa. Pored toga eksperimenti mogu biti izvršeni na lokalnim radnim stanicama.



Slika 7 Struktura sistema prepoznavanja govora – Spear

[7] http://pypi.python.org/pypi/bob.spear
[8] http://www.idiap.ch/software/bob/buildbot/waterfall
[9] http://pypi.python.org/pypi/facereclib

4.1.1. PROCESIRANJE

Ovaj alat ima implementirane dve tehnike za detekciju govornih aktivnosti (SAD – speach activity detection).

Prva tehnika je jednostavna bez nadzora na energiji zasnovana tehnika (unsupervised energybased SAD) u okviru koje se okvirni nivoi energije sabiraju, normalizuju i onda klasifikuju u dve klase. Klasa sa najvišom srednjom vrednošću se smatra govorom i odgovarajući govorni segmenti se čuvaju pre nego što se poravnaju.

Druga tehnika se zasniva na kombinaciji energije i njena modulacija je oko 4Hz.

Jednostavan i prilagodljiv thresholding je primenjen na obe funkcionalnosti za otklanjanje delova koji ne spadaju pod govor. Ova tehnika se pokazala kao dobra u mobilnom okruženju. Pored toga, ovaj alat podržava i upotrebu eksternih ili ručno kreiranih tehnika za detekciju govornih aktivnosti.

Spear pruža efikasnu implementaciju spectral i cepstral funkcija (MFCC – *eng. mel-frequency cepstrum coefficient* i LFCC – *eng. linear frequency cepstral coefficients*), koje opciono mogu kombinovane sa njihovim prvim i drugim derivatima.
Cepstral srednja vrednost i Variance tehnika normalizacije su, takođe, integrisane u alat. Pored toga, alat pruža interfejs za čitanje HTK i Spro funkcionalnosti.

Gaussian mixture model (GMM), združena faktorska analiza (joint factor analysis), inter-session variability, total variability (i-vectors) state-of-the-art tehnike modelovanja su uključene u ovaj alat. One se oslanjaju na efikasnu C++ implementaciju koja je dostupna u Bob-u.

GMM sistem uključuje univerzalni pozadinski model (UBM) obuke korišćenjem maksimalne verovatnoće (ML – maximum-likelihood). Jednostavna i paralelna implementacija treninga su podržane.

Upis modela se izvršava korišćenjem maximum-posteriori (MAP) prilagođavanja.
Spajanje rezultata i posmatranog modela se vrši uz pomoć log faktora rizika (LLR). Linearna aproksimacija LLR-a je, takođe, podržana sistemom. Pored toga ZTnorm se može opciono koristiti za rezultate normalizacije.

Ostale tri tehnike modelovanja su ugrađene na vrhu GMM modelovanja. Jednom kada je UBM trening postignut statistike nultog, prvog i drugog reda se izračunavaju pri svakom govornom iskazu.

Za JFA sistem statistika izgovaranja pripada trening skupu koji se koristi za procenu matrice sopstvenog glasa (eigenvoice – V), matrice vlastitog kanala (eigenchannel – U) i matrice preostale buke (D). Povezivanje govornih modela i govornih izjava se obavlja korišćenjem linearnog bodovanja aproksimacija (FOOTNOTE 24). Sistem ISV je sličan JFA sistemu, sa jedinom razlikom što se ne procenjuje matrica sopstvenog glasa (V). Konačno, sistem i-vectors uključuje računanje matrice totalnog varijabiliteta (total variability matrix T), za čije procenjivanje se koristi isti algoritam kao i za trening V. U svrhu tih procena dostupne su i jednostavna i paralelna implementacija. Nisko-dimenzionalni i vektori se nakon toga ekstrahuju iz svake govorne izjave.

Izbeljivanje (whitening), dužina normalizacije (length normalization), diskriminaciona linearna analiza (linear discriminant anaysis LDA) i normalizacije kovarijanse (within-class covariance normalization –WCCN) su uključene u alat, i imaju za cilj mapiranje u više adekvatnih prostora. Upoređivanje između rezultata testa i posmatranog govora se vrši pomoću fast scoring-a zasnovanog na udaljenosti izmedju i-vektora ili probabilističkom diskretnom linearnom analizom (PLDA). Ovaj alat koristi efikasnu i skalabilnu implementaciju PLDA.

U poslednjih nekoliko kampanja ocene prepoznavanja govora kao što su NIST SRE i Mobio, utvrđeno je da je fuzija značajno poboljšala performanse prepoznavanja. Dakle, score fusion strategija zasnovana na logičkoj regresiji je, takođe, integrisana u alat.

4.1.2. ODLUKA I EVALUACIJA

Mere za evaluaciju, kao što su jednaki procenat greške (Equal error rate, EER), half-total procenat greške (HTER), funkcija minimalnih troškova odluke (minDCF), utvrđivanje greške kompromisa (detection error trade-off DET), prijemnik operativnih karatkeristika (receiver operating characteristic ROC), kriva očekivanh performansi (EPC), log-verovatnoća odnosa troškova (CLLR) i minimalna log-verovatnoća odnosa troškova (min-CLLR) su implementirani u Bob i

dostupni u Spear alatu. To omogućava podešavanje hiper parametara i procenu sistema veoma lako i unapred.

4.1.3. BAZA PODATAKA

Za poboljšanje ponovljivosti i uporedivosti naučnih publikacija, Spear koristi nekoliko baza podataka sa dobro definisanim protokolima.
Protokoli kao što su NIST SRE 2012, MOBIO, BANCA, TIMIT i Voxforge 12 su uključeni u ovaj alat. Razvojni set NIST SRE 2012 protocol13 je pripremljen sa I4U partnerima tokom njihovog učešća u evaluaciji. Voxforge je open-source baza podataka za koju je kreiran open-soruce protocol14. On se koristi kao primer koji omogućava istraživačima da slobodno prolaze i testiraju alat.
Konačno, eksperimenti u lokalnim bazama podataka, koje definiše korisnik mogu biti lako sprovedeni.

4.1.4. EKSPERIMENT

Kao dokaz koncepta prikazaćemo eskperimentalno poređenje između GMM, ISV i i-vektora i njihovu fuziju na mobile-0 protokolu MOBIO baze podataka (E. Khoury, 2013).
U ovim eksperimentima GMM-ovi su sačinjeni od 512 Gaussian komponenti. Za ISV, rang podprostora U je podešen na 50, dok je za i-vektore rang T podešen na 40015. U svrhu sprovođenja reprodukcionog istraživanja ovaj alat može biti proširen algoritimima koji mogu biti integresani u bilo kojoj fazi istraživanja.

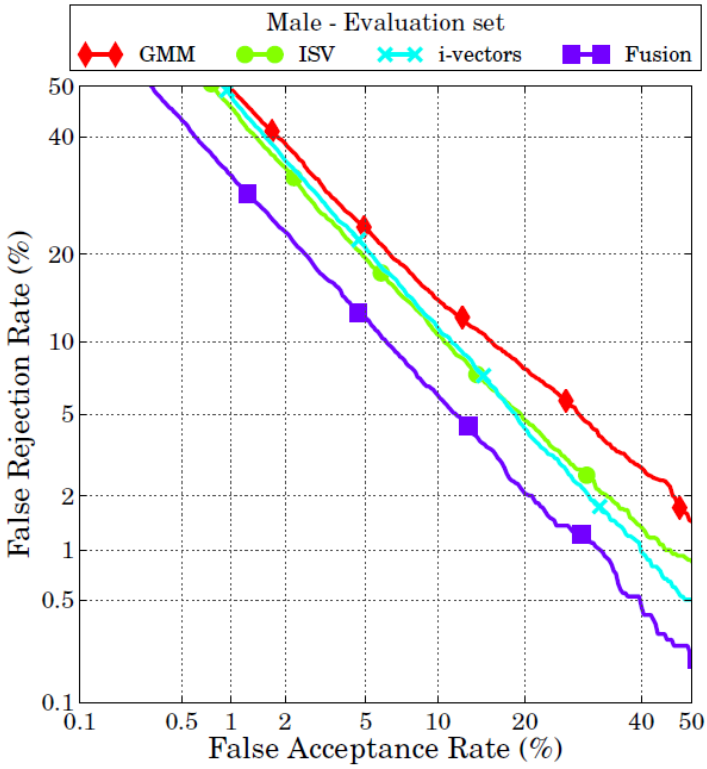
U narednoj tabeli prikazani su rezultati četiri sistema i to EER na Dev set, HTER na EVAL set i min-CLLR na oba (DEV set i EVAL set) i to za muške i ženske govornike.

Tabela 1 Rezultati (DEV set i EVAL set)[10]

	Set	Measure	GMM	ISV	i-vectors	Fusion
Male	DEV	EER	13.41	10.40	11.31	7.31
		min-CLLR	0.4509	0.3708	0.3795	0.2610
	EVAL	HTER	12.12	10.36	11.11	7.89
		min-CLLR	0.4189	0.3621	0.3621	0.2769
Female	DEV	EER	17.94	12.22	12.59	9.21
		min-CLLR	0.5693	0.4214	0.4182	0.3197
	EVAL	HTER	17.68	16.23	17.36	14.65
		min-CLLR	0.5464	0.5156	0.5641	0.4813

Možemo primetiti da su, baš kao što se navodi i u literaturi, ISV i i-vektorski sistemi su tačniji od GMM osnovne line. U tabeli vidimo da je ISV malo bolji od i-vektora na bazi podataka koja je korišćena za primer. Ovo se može objasniti nedostatkom podataka potrebnih za trening različite faze i-vektora. Poslednja kolona tabele, takođe, pokazuje da kombinovanje tri samostalna sistema znatno povećava sistem performanse: HTER – 7,9% i 14,7% za muške i ženske govornike, respektivno. Slični trendovi su prikazani i pomoću DET krivih.

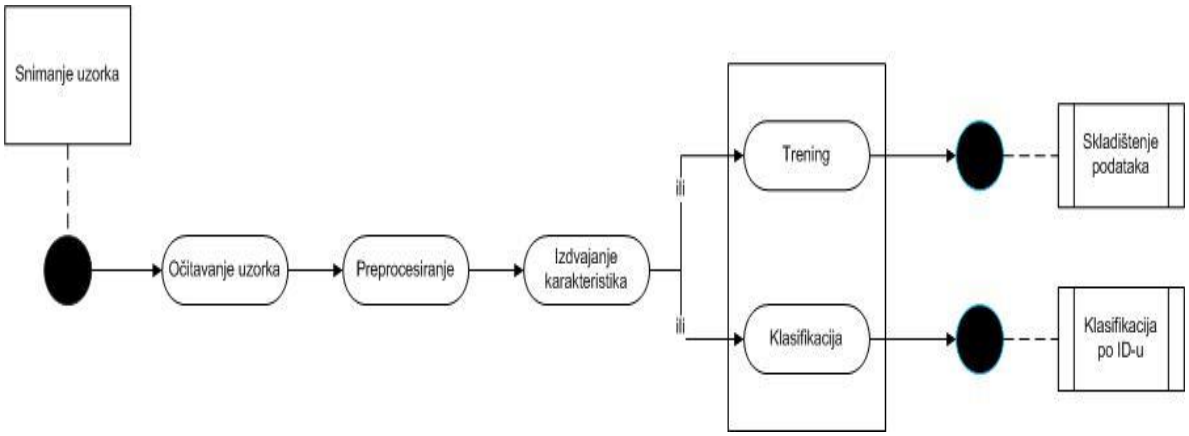
[10] http://pypi.python.org/pypi/bob.spear



Slika 8 Det kriva za prikazanu evaluaciju Mobio baze podataka i protokola MOBILE-0 (Mondal, Bours, & Idrus, 2013)

4.2. MARF

Marf je open-source platforma koja predstavlja kolekciju glasa, zvuka, govora, teksta i obrade prirodnih jezika napisanih u Javi gde se omogućava lakše dodavanje novih algoritama. Obrada prirodnog jezika je oblast veštačke inteligencije koja se bavi proučavanjem problema automatskog proizvođenja i razumevanja prirodnih ljudskih jezika. Današnji Marf (slika 9) obuhvata uzorak za prepoznavanje (snimanje, pretprocesiranje , izdvajanje karakteristika, trening i klasifikacija), gde su prikazani razni algoritmi, obrada prirodnog jezika, kao i distribuirana verzija (Mokhov, 2006).



Slika 9 Struktura sistema prepoznavanja glasa – MARF

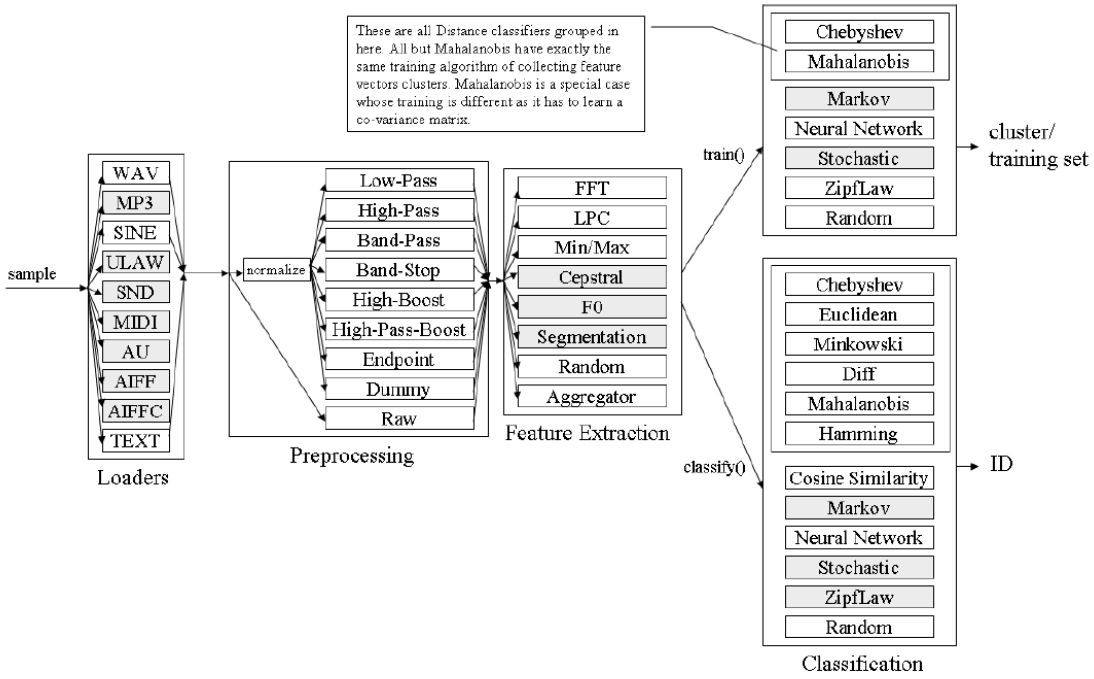
Marf je open-source istraživačka platforma i kolekcija glasova/zvukova/govora/teksta i algoritama za obradu prirodnog jezika (NLP) napisanih u Javi i organizovanih u modularnom i proširivom frameworku koji omogućava dodavanje novih algoritama. MARF može da se distribuira preko mreže, ali može da se koristi i kao biblioteka u aplikacijama ili da se koristi kao izvor za učenje i proširenje.[11]

Nekoliko primera aplikacija je sadržano u samom paketu kako bi se prikazao okvir rada, takođe je sadržan i priručnik za upotrebu koji detaljno objašnjava način upotrebe, kao i API reference u javadoc formatu iz čega zaključujemo da je projekat dobro dokumentovan. MARF i njegove aplikacije su objavljeni pod BSD licencom i hostovani na SourceForge.net i sve projektne strane mogu biti pronađene tamo.

MARF predstavlja kolekciju API paterna generalne namene za prepoznavanje i njihovu realizaciju. U početku, MARF je razvijan da pokaže kako funkcionišu implementirani algoritmi i istovremeno da obezbedi istraživačima platformu za testiranje novih algoritama, jednih protiv drugih u različitim performansama (npr. Run-time i potrošnja memorije) korišćenjem zajedničkih platformi. MARF i njegovi subframework-ovi i aplikacija originalno su se razvili iz koncepta audio prepoznavanja, ali nisu bili ograničeni na tu temu u svojoj opštosti kao ni na implementirane algoritme.

[11] http://marf.sourceforge.net/

Marf pokriva paterne prepoznavanja pipeline (učitavanje, procesiranje, ekstrakcija, trening i klasifikacija), kao što je prikazano na slici, razne algoritme, PureData plugin, procesiranje prirodnog jezika i distribuirane verzije.



Slika 10 Marf pattern prepoznavanja – pipeline[12]

4.2.1. PIPELINE

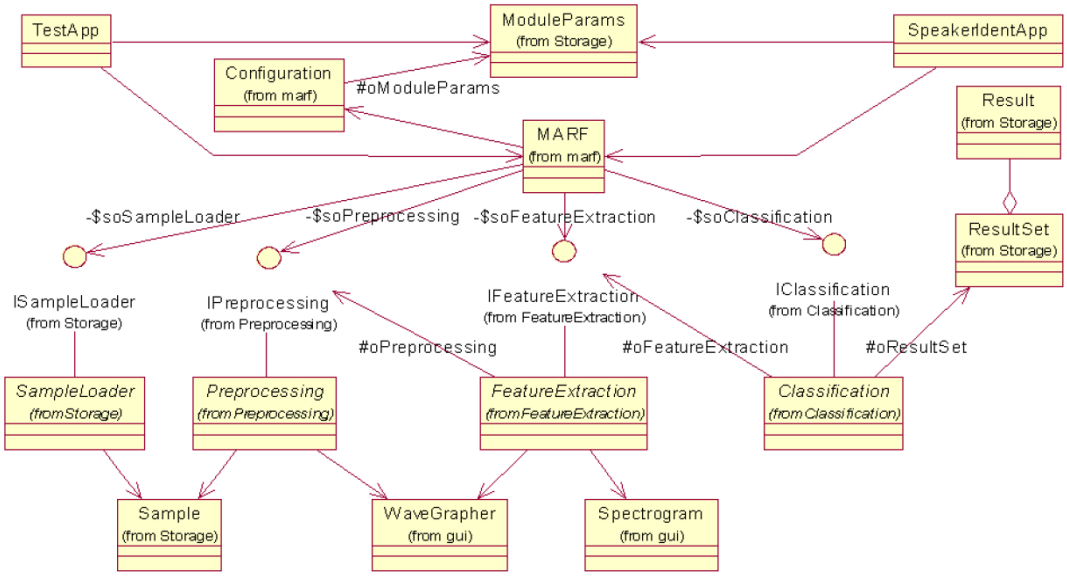
Konceptualno pattern prepoznavanja pipeline je prikazan na slici i objašnjava protok podataka izmedju faza. Generalno, ceo patern prepoznavanja počinje učitavanjem sample-a (npr. Neki audio zapis u formi talasa), nastavlja se njegovim procesiranjem, zatim raspakivanjem najistaknutijih karakteristika i na kraju trening sistema tako da sistem ili uči novi skup karakteristika subjekta ili samo klasifikuje i identifikuje šta je, odnosno ko je subjekat. Ishod treninga je bilo koja kolekcija nekog oblika vektora ili njihovo sredstvo srednjeg klastera, koje se snima za svaki subjekat. Rezultat je 32bitni jedinstveni intidžer koji obično ukazuje ko/šta je subjekat, odnosno šta sistem misli da je subjekat.

[12] http://marf.sourceforge.net/report/auto/report.pdf

4.2.2. APLIKACIJA

Postoji velik broj aplikacija koje mogu biti napisane na osnovu MARF-a u njegovom današnjem obliku. U okviru Marf-a je MARF's CVS repozitorijum, postoje četiri glavna i mnogo manjih testova i demo aplikacija koje se nalaze u paketu sa MARF-om, u zavisnosti od objavljene verzije.

Četiri osnovne aplikacije uključuju Text-independent speaker identification, language identification, probabilistic parsing i Zipf's law analysis. The former je najrazvijanija aplikacija koja je koristila varijacije testova specifičnih za MARF, uključuje iscrpno testiranje inplementiranih algoritama sa glavnim ciljem da ilustruje najbolju moguću konfiguraciju pretprocesiranja, funkciju izvođenja i klasifikuje algoritme koji daju najbolju tačnost i tako pomognu istraživačima da se fokusiraju na algoritme koji su potrebni za specifičan domen njihovog problema. Sve to može biti prikazano pomoću UML dijagrama.



Slika 11 Generalna upotreba MARF-a i implementacija njegovih aplikacija[13]

4.2.3. ALGORITMI

Na vrhu frameworka se nalazi sam API, MARF ima nekoliko stvarnih implementacija brojnih algoritama u svrhu prikaza svojih mogućnosti u različitim fazama cevovoda (pipeline) i pratećih modula. Lista nekih od implementiranih algoritama obuhvata:

[13] http://marf.sourceforge.net/report/auto/report.pdf

- Linearno prediktivno kodiranje (LPC) – koristi se za izvođenje osobina
- Fast Fourier transformacija (FFT) – koristi FFT zasnovane filtere kao i ekstrakciju svojstava.
- Broj na distanci zasnovanih klasifikatora (Euclidean, Chebyshev, Minkowski, Mahalanobis, Diff (interno razvijeni sa projektom po ponašanju slični UNIX/Linux diff alatu) i Hamming).
- Mera kosinusa sličnosti, koja je istražena u dubinu i doprinosi najboljoj tačnosti u ovom poslu za mnoge konfiguracije.
- Neuronske mreže – koriste se u klasifikaciji
- Zipfov zakon zasnovan na klasifikatoru.
- Kontinuirana frakcija proširenja (CFE)- zasnovana na filterima.
- Određeni broj na matematici zasnovanih funkcionalnosti - na primer, matrica i vektor obrade, uključujući kompleksne matrice brojeva i vektorske operacije i statističke procene.

Budućnost MARF-a biće usmerena uglavnom na poboljšanje kvaliteta postojećeg koda, na dodatno pojednostavljenje i implentaciju nedostajućih bitnih funkcionalnosti.

Tačnost MARF sistema može da se zasniva na tehnikama pre-procesiranja koje se koriste za odabir karakteristika koje proizvodi glas i za identifikaciju govornika (Jadhav & Dharwadkar, 2018). U ovom primeru se tačka kontinuiranog kvaliteta eliminiše u procesu normalizacije. Cepstralni koeficijent frekvencije Mel-skale je jedan od metoda za preuzimanje karakteristika iz talasne datoteke izgovorenih rečenica. Takođe, ovde se koristi model govornika baziran na Gausovim mešavinama gde su rađen eksperimenti na MARF-u radi povećanja procene ishoda. Tu je predstavljena eliminacija krajnje tačke u medijumu Gausove selekcije za MFCC.

Još jedno istraživanje koje je usmereno na MARF platformu je istraživanje u kombinaciji sa Illimitable Space Sistem (ISS). ISS je interaktivni konfigurabilni alat u realnom vremenu koji umetnici koriste za kreiranje interaktivnih vizuelnih efekata u pozorišnim predstavama i u dokumentarnim filmovima putem korisničkih unosa kao što su gestovi i glas. U ovom radu, pored postojećih mogućnosti za obradu vizuelnih i geometrijskih podataka zasnovanih na pokretu, opisujemo se obrada zvuka uz pomoć MARF-a. Sa ovim dodatnim modulima, ISS može pomoći interaktivnom kreiranju performansi koje koristi vizuelno praćenje i obradu signala kako bi se

kreirale animacije u obliku čoveka koje se mogu pratiti u realnom vremenu (Song, Mokhov, Song, & Mudur, 2018).

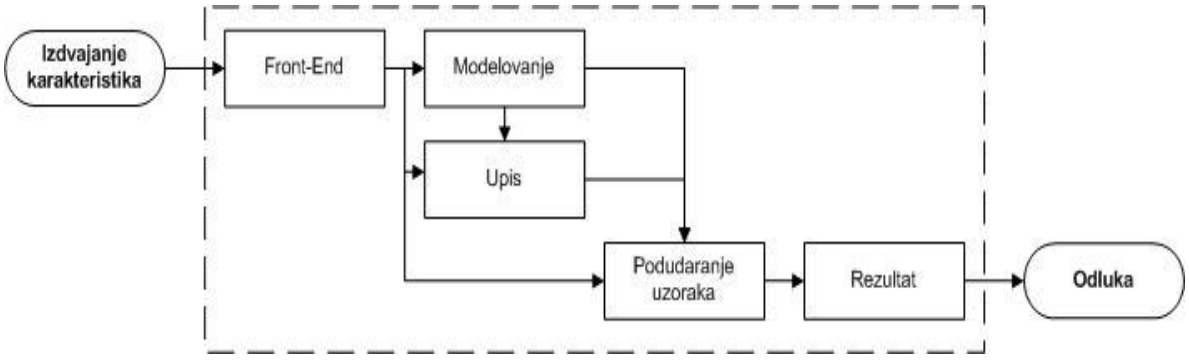
4.3. ALIZE

ALIZE[14] je open source platforma razvijena za prepoznavanje glasa 2005. godine. Najnovija verzija (3.0) uključuje state-of-the-art metode kao što su Joint Factor Analysis i I-vector modeling i Probabilistic linear discriminant analize. C++ multiplatformska implementacija Alize je dizajnirana da rukuje većom količinom podataka potrebnih za glasovnu i jezičku detekciju i da olakša razvoj state-of-the-art sistema.

ALIZE je pokrenut od strane Univerziteta Avignon-LIA 2004. godine. Od tada, mnogo istraživačkih institucija i kompanija je doprinelo razvoju projekta. Ova saradnja je doprinela velikom broju alata koji se mogu pronaći na ALIZE website-u (http://alize.univ-avgnon.fr).[15]

Softverska arhitektura ALIZE platforme se zasniva na UML modelovanju i strogim konvencijama vezanim za kod, a sve u cilju olakšavanja zajedničkog razvoja i održavanja koda. Kao open source i cross-platforma testova omogućava njegovim korisnicima da brzo pokrenu test regresije u cilju povećanja pouzdanosti novih verzija i radi lakšeg održavanja. Test slučajevi uključuju nizak nivo unit testova u osnovi, najvažnije algoritamske klase, kao i integracione testove. Doxygen dokumentacija je dostupna online i može biti kompajlirana iz samo source koda.
Kenny et al. (Kenny, 2004) su opisali implementaciju za prepoznavanje govornika gde su primenili Faktorsku analizu da bi objasnili kako određeni faktori utiču na sam pristup prepoznavanja glasa.

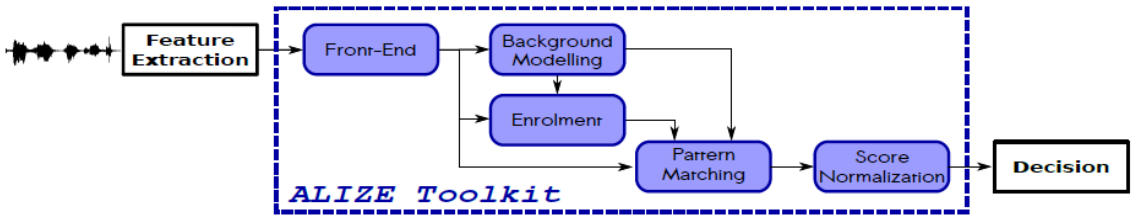
[14] http://alize.univ-avignon.fr/
[15] http://www.signalprocessingsociety.org/technical-committees/list/sl-tc/spl-nl/2013-05/ALIZE/



Slika 12 Struktura sistema prepoznavanja glasa – ALIZE

4.3.1. STRUKTURA SISTEMA

Platforma uključuje Visual Studio rešenja i autotools za kompajliranje u okviru Windows i UNIX platformi. Veliki deo LIA RAL funkcija koriste paralelno procesiranje za brzu multi-thread implementaciju zasnovanu na standardnim POSIX bibliotekama koje su u potpunosti kompatibilne sa većinom sličnih platformi. LIA RAL biblioteka se može povezati sa dobro poznatom Lapack3 bibliotekom koja obezbeđuje veliku preciznost u manipulaciji matricama. Svi kodovi su dostupni u okviru LPGL licence koja sadži minimalne restrikcije u pogledu redistribucije softvera.



Slika 13 Generalna struktrua sistema za verifikaciju govora[16]

Šema prikazuje arhitekturu alata engine-a za prepoznavanje govora. LIA_RAL obezbeđuje izvršne datoteke koje se mogu koristiti kako bi se ostvarile različite funkcije, što je prikazano na ovom dijagramu.

U prepoznavanju govora, modul za upis generiše statistički model iz jedne ili više sekvenci. Iako je moguće generisati jedan model za svaku sesiju snimanja, u zavisnosti od arhitekture sistema, uobičajeno je da se razmotri jedan model za predstavljanje govornika.

[16] http://www.signalprocessingsociety.org/technical-committees/list/sl-tc/spl-nl/2013-05/ALIZE/

State-of-the-art speaker recognition engines su uglavnom vezane za tri tipa govornih modela. Iako su ta tri modela povezana sa Gaussian mixture modelom (GMM) razlikuju se po prvom tipu modela koji se eksplicitno koristi za GMM i druga dva tipa koja reprezentuju govornika ili sesije kao vektore fiksne dužine izvedene iz GMM. Ovi vektori mogu biti super-vektor ili više kompaktna reprezentacija poznata kao i-vektor. Svaki od tri modela mogu se dobiti iz odgovarajućeg izvršnog alata LIA RAL paketa.

Nedostatak predstavlja to što varijabilnost sesija može da zavisi od kanala i buke i jedno od glavnih pitanja jeste prepoznavanje govornika. Alize obuhvata najćešća rešenja za ovaj izazov što će biti opisano.

4.3.2. TRAINTARGET

TrainTarget generiše GMM model is jedne ili više sekvenci. U GMM se može prilagoditi univerzalna pozadina modela (UBM) *M* korišćenjem maksimum, a posteriori (MAP) kriterijuma ili maksimalne verodostojnosti linearne regresije (MLLR). Buka i kanal robustnosti prikaza mogu se dobiti primenom Factor analize (FA). Faktorska analiza za prepoznavanje govornika je uvedena i pretpostavlja da varijabilnost proizilazi iz toga što između govornika i kanala postoji međuprostor. Dve različite FA analize su dostupne u okviru ALIZE. Jedna, koja je poznata kao Joint Factor Analysis, gde je super-vektor $m_{(s,n)}$ od sesije i govornika zavisnog GMM je suma tri uslova data u primeru.

$$\mathbf{m}_{(s,n)} = \mathcal{M} + \mathbf{V}\boldsymbol{y}_{(s)} + \mathbf{U}\boldsymbol{x}_{(s,n)} + \mathbf{D}\boldsymbol{z}_{(s)}$$

U ovoj formuli, *V* i *U* su „faktori učitane matrice“, *D* je dijagonala MAP matrice dok su $y_{(s)}$ i $x_{(s,n)}$ respektivno nazvani govornik i faktori kanala. $y_{(s)}$, $x_{(s,n)}$ i $z_{(s)}$ su nezavisni i imaju normalnu standardnu distribuciju. Pojednostavljena verzija ovog modela je obično poznata kao EigenChannel ili Latent Factor Analysis i takođe je dostupna u TrainTarget-u. U ovoj verziji, pojednostavljena generativna jednačina izgleda ovako:

$$\mathbf{m}_{(s,n)} = \mathcal{M} + \mathbf{U}\boldsymbol{x}_{(s,n)} + \mathbf{D}\boldsymbol{z}_{(s)}$$

4.3.3. MODETOSV

ModeToSV izvodi super vektor iz GMM-a. Super vektor predstavlja govornika u visoko dimenzionalnom prostoru što je postalo popularno sa razvojem Support vector mašina za prepoznavanje govora (SVM). LibSVM biblioteka je integrisana u ALIZE.

Nuisance Attribute Projection (NAP) ima za cilj da ublaži varijabilnost kanala u prostoru suprvektora odbijanjem medjuprostora koji sadrži najveći deo varijabilnosti kanala (efekat smetnje). Najjednostavniji pristup sastoji se od procene sesije kovarijanse matrice *W* seta govora i izračunavanja projekcije $\hat{m}_{(s,n)}$ super-vektora $m_{(s,n)}$:

$$\hat{\mathbf{m}}_{(s,n)} = (I - SS^t)\mathbf{m}_{(s,n)}$$

Gde S sadrži prvi eigenvectors koji proizilazi iz jedinstvene vrednosti dekompozicije *W*.

4.3.4. IVEXTRACTOR

IvExtractor izvodi nisko-dimenzionalni i-vektor iz sekvenci obeležja vektora. I-vektori se genrišu na osnovu:

$$\mathbf{m}_{(s,n)} = \mathcal{M} + \mathbf{T}w_{(s,n)}$$

Gde je T niska pravougaona matrica koja se naziva Matrica totalne varijabilnosti i i-vektor $w_{(s,n)}$ je probabilistička projekcija super vektora $m_{(s,n)}$ na ukupnu varijabilnost prostora definisanu kolonama T.

Mnoge tehnike normalizacije se koriste u svrhu smanjenja sesije varijabilnosti u ukupnoj varijabilnosti. IVNorm se može koristiti da se izvrši normalizacija zasnovana na Eigen Factor Radila (EFR) metodu.

4.3.5. POZADINSKO MODELOVANJE

Prepoznavanje govornika je data-driven tehnologija i svi pristupi implementirani u ALIZE se u pozadini oslanjaju na velike količine podataka. Procena komponente znanja je računski intenzivna. U tu svrhu su razvijeni efikasni alati koji služe za optimizaciju i pojednostavljenje razvojnih napora. UBM može biti korišćen efikasno pomoću TrainWorld-a koji koristi nasumično odabrane funkcije da ubrza iterativni proces učenja zasnovan na EM algoritmu.

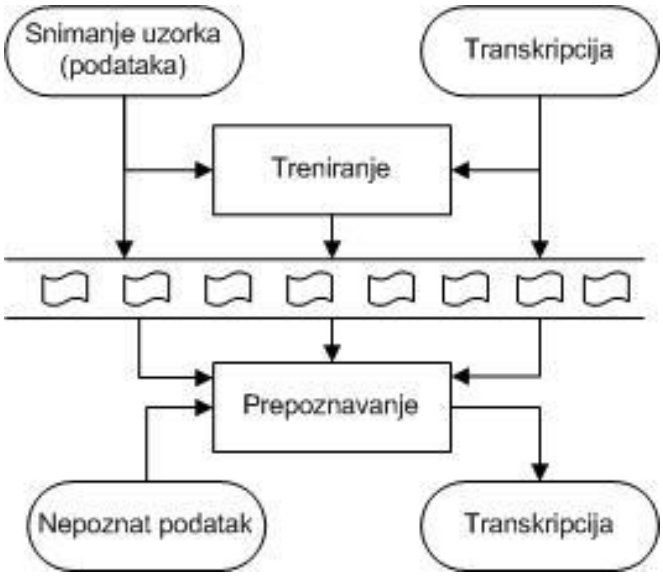
Meta Parametri JFA i LFA modela mogu biti korišćeni uz pomoć EigenVoice, EigenChannel i EstimateDmatrix.

4.4. HTK

HTK (HMM toolkit) je alat koji se u opštem slučaju koristi za izgradnju i modelovanje skrivenih Markovljevih modela. Glavnu primenu nalazi u izgradnji sistema za prepoznavanje govora.

Osnovni princip rada tog alata se može objasniti u nekoliko koraka (Steve Young, 2001) – postoje dva osnovna procesa koji se odvijaju sledno: trening i prepoznavanje. Pre nego što se pokrenu oba procesa, bitno je dobro pripremiti sve podatke potrebne za prepoznavanje. Koraci su sledeći (Kolak A. , 2009):

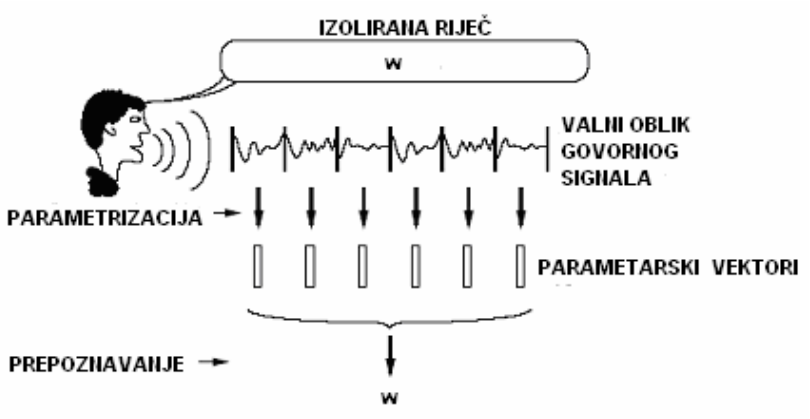
- *Priprema podataka* – obavlja se snimanje materijala za fazu treninga i testa, te se kreiraju transkripcije (ako govor nije označen), definiše se vokabular koji je deo procesa prepoznavanja. Na osnovu vokabulara izrađuje se rečnik transkripcija na temelju vokabulara materijala za trening i test, i na kraju se obavlja rekonstruisanje govornog signala (npr. wav u mfc)
- *Treniranje* – definišu se modeli, odnosno vrsta modela koja zavisi od načina treniranja (metoda izolovanih reči ili metoda slednih reči). Na osnovu uzoraka iz faze treninga obavlja se estimacija parametara na tim modelima.
- *Prepoznavanje* – Kada se završi izgradnja i procena modela kako inicijalno zamislilo, dolazi do neidentifikovanih izgovora koji se koristi u fazi testa, kao i dobijanje konačnih razultata posredstvom pravila utvrđenih gramatikom.



Slika 14 Struktura sistema za prepoznavanje glasa – HTK

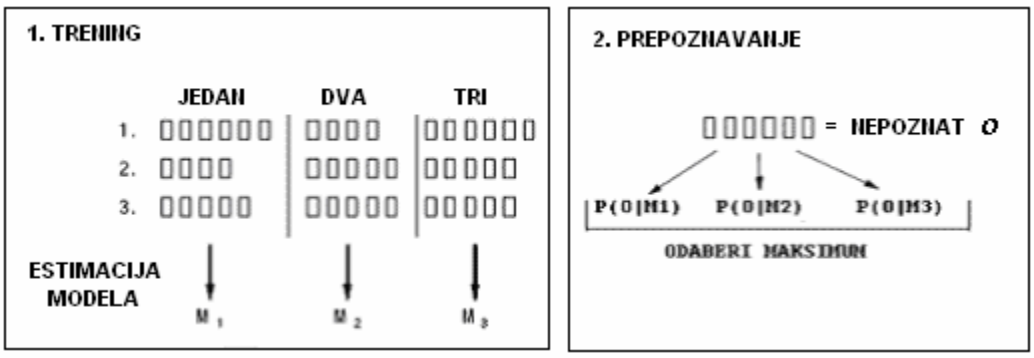
4.4.1. METODA PREPOZNAVANJA IZOLOVANIH REČI

Kod načina prepoznavanja metodom izolovanih reči (Kolak, 2009), svaka reč iz baze za trening ima svoj model. Kako jedan govorni iskaz predstavlja jednu reč (u slučaju da su sklonjene tišine), nema potrebe za ručnim označavanjem materijala, nego se koriste pripadne transkripcije. Ovakvi modeli se najčešće inicijalizuju pomoću alata Hinit, a re-estimiraju pomoću HRest. U slučaju kad nema dovoljno podataka za trening, bolji rezultati se postižu modelima kojima su sve varijanse izjednačene s globalnom, a to se postiže prethodnom inicijalizacijom pomoću HCompV. Prepoznavanje se vrši pomoću HVite.



Slika 15 Prepoznavanje izolovanih reči[17]

Jednostavan primer prepoznavanja izolovanih reči je sledeći. Na primer, imamo vokabular od tri reči: jedan, dva i tri. Za trening se modeli treniraju, svakog bude u više primeraka. Kad je u pitanju neka nepoznata reč, onda se računa verodostojnost da je ta reč generisana svakim od modela i kao rezultat se uzima model sa najvećom verodostojnošću, što vidimo na slici 16 (Kolak, 2009).



Slika 16 Primer prepoznavanja govora na izolovanim rečima[18]

4.4.2. METODA PREPOZNAVANJA SLEDNOG GOVORA

Kod ove metode, modeli se definišu na temelju reči (ako se radi o vezanom govoru), ili fonemima (ako se radi o kontinuiranom govoru). Zavisno od količine materijala, može se raditi sa ručno označenim materijalima (ako ih je manje), ili se radi sa neoznačenim materijalom (koji ima pripadne transkripcije). Prvi način koristi inicijalizaciju pomoću Hinit i HRest, kao i metoda

[17] https://bib.irb.hr/datoteka/451087.Zavrsni_Kolak_Antonio_0036427170.pdf
[18] https://bib.irb.hr/datoteka/451087.Zavrsni_Kolak_Antonio_0036427170.pdf

izolovanih reči, dok se u drugom načinu koristi flat start, pomoću HCompV. Drugi način sa kontinuiranim govorom je najčešći slučaj kod sistema za prepoznavanje govora, jer omogućava neograničeni vokabular za trening na temelju ograničenog broja modela (fonemi). Re-estimacija se uvek vrši funkcijom HERest, dok se prepoznavanje radi sa HVite, koja sad koristi token passing algoritam.

4.4.3. HTK FUNKCIJE

Kao i svaki alat, tako i HTK alat ima svoje funkcije koje možemo da pozivamo iz određenih programa. Obično se koriste sledeće:

- **Hcopy** - koristi se za konverziju ulaznog zvučnog signala u neki od parametriziranih oblika. Važno je u konfiguracijskoj datoteci odabrati stavke za parametrizaciju (ako se radi mfc parametrizacija) (Kolak A. , 2009):
  SOURCEFORMAT → format snimljenog materijala

  TARGETKIND → ciljani parametri

  TARGETRATE → period okvira (HTK koristi jedinice od100ns)

  WINDOWSIZE → veličina okivra

  USEHAMMING → (T/F) koristi li FFT analiza koristi Hammingov otvor

  PREEMCOEF → koeficijent prednaglašavanja signala

  NUMCHANS → broj kanala u filtarskoj banci

  NUMCEPS → broj MFCC koeficijenata

- **HcompV** - računa globalnu srednju vrednost i varijansu iz ulaznih podataka za trening. Prema "flat start" trening metodi, ***HCompV*** služi za inicijalizaciju u kojoj se svakom stanju u svakom modelu monofona dodeljuju globalne srednje vrednosti i varijanse. Takođe, moguće je odrediti donju granicu iznosa varijanse – što je vrlo korisno u slučaju kad se veliki set modela trenira na osnovu ograničene količine podataka za trening. U protivnom bi iste, zbog nedostatka podataka, mogle biti pogrešno estimirane.

- **HInit -** Kada je pripremljen skup različitih prototipa HMM modela, može se pristupiti procesu inicijalizacije korišćenjem alata *HInit*. Osnovni princip *HInit* alata bazira se na konceptu HMM-a kao generatora vektora obeležja. Svaki uzorak za obuku sistema može se posmatrati kao izlazna sekvenca HMM čiji parametri treba da se izračunaju. Prema tome, ako je poznato koja stanja su generisala svaki vektor obeležja u govornoj bazi za obuku, tada se nepoznate srednje vrednosti i varijanse posmatranog stanja HMM-a mogu izračunati na osnovu svih vektora obeležja povezanih sa datim stanjem.

- **Hrest -** Ovo je alat veoma sličan alatu *HInit*, s tim što on zahteva da su ulazne definicije HMM-ova već inicijalizovane, a umjesto Viterbijevog algoritma koristi tzv. *Baum-Welch* re-estimacioni postupak. On uključuje pronalaženje verovatnoće da se model nalazi u određenom stanju tokom trajanja nekog vremenskog intervala primenom *Forward-Backward* algoritma. Prema tome, dok Viterbijev algoritam donosi grubu odluku od strane kog stanja je generisan koji vektor obeležja. Baum-Welch algoritam omogućava da jedan vektor bude pridružen nekolicini različitih stanja, ali sa različitom verovatnoćom. Ovaj alat se koristi samo za obuku monofona (fonemi nezavisni od konteksta u kom se nalaze), jer kao i *Hinit* zahteva da mu se kao ulazni podaci proslede prototip i ime modela.

- **HERest** - provodi jednostruku re-estimaciju parametara celog seta modela pomoću Baum-Welch algoritma. Za svaku ulaznu rečenicu sa pripadajućom transkripcijom ***HERest*** ulančava one modele fonema koji odgovaraju fonemima na listi i formira jedan komponirani model. Nad ovakvim modelom provodi forwardbackward algoritam i skuplja statistike o okupaciji stanja, srednjim vrednostima i varijansama za svaki model u lancu (Kolak A. , 2009). Nakon što su na ovaj način obrađeni svi uzorci, ukupne statistike se koriste za re-estimaciju parametara modela. Ovakvu re-estimaciju je dovoljno sprovesti 2 do 5 puta. Višestruka ponavljanja re-estimacije osim što mogu dugo trajati, mogu i rezultovati sistemom koji je i toliko dobro utreniran na vokabular za trening da se ne može prilagoditi na nove reči pri testiranju.

  Važno je napomenuti da su transkripcije potrebne samo da se definiše tačan niz fonema u svakom uzorku, informacija o granici među fonemima nije potrebna.

- **Hled** - služi kao editor datoteka sa transkripcijom. Funkcija učitava niz naredbi iz edit skripte (sa .led ekstenzijom), i datoteku s jednom ili više labela, te ispisuje novu datoteku s labelom uređenu prema naredbama iz edit skripte.
  U edit skripti svaka naredba mora biti zapisana u svom redu, a naredbe se dele na dve grupe: naredbe koje se primenjuju na zasebne labele, naredbe koje se primenjuju na celi set labele, tj. na celu datoteku s transkripcijom – MLF

- **Hhed** - učitava set modela i edit skriptu (nastavak .hed) s nizom naredbi na osnovu kojih provodi transformacije nad nekim modelom ili setom modela (Kolak A. , 2009).
  Najčešće transformacije koje provodi ***HHEd*** su: Kopiranje modela da bi se dobio kontekstno zavisni set, povezivanje (deljenje) parametara, raspoređivanje u razrede (klastering), dodavanje ili uklanjanje prelaza među stanjima

- **HParse -** služi za kreiranje strukture (mreže), na osnovu gramatike u kojoj su specifikovane sekvence reči koje su dozvoljene. Mreža na nivou reči reprezentuje gramatiku koja je napisana u formi Backus-Naur notacije gde se prepoznavanje sastoji od skupa čvorova povezanih granama. Svaki čvor je ili HMM model ili kraj reči. Svaki HMM model za sebe predstavlja mrežu koja se sastoji od stanja povezanih granama, pa se samim tim mreža za prepoznavanje sastoji od stanja HMM modela povezanih granama prelaza.

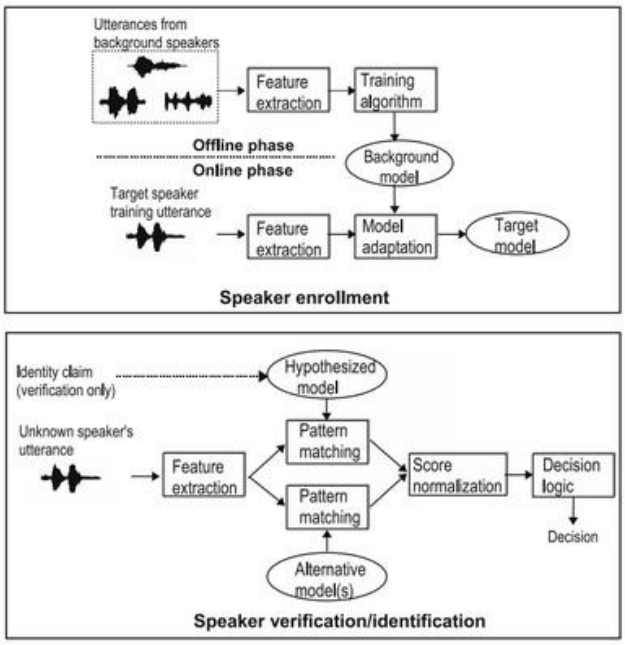
- **HResult** - Alat *HResult* porede i transkripciju dobijenu prepoznavanjem sa originalnom transkripcijom generiše raznovrsne statistike vezane za performanse dobijenog ASR sistema.

- **HVite -** Nakon što je pripremljena mreža koja definiše reči koje se očekuju i kako se svaka od tih reči čita može se pristupiti prepoznavanju govora. Zadatak prepoznavanja govora je da nađe najverodostojniju putanju kroz mrežu za dat nepoznat govorni fajl i HMM modele.

4.5. SUMARNI PRIKAZ FUNKCIONALNOSTI PROGRAMSKIH REŠENJA

Treba spomenuti da je ovaj rad rađen zbog aktuelnosti biometrijske tehnologije gde svaka od njih koristi različite biometrijske senzore i različite algoritme za uparivanje biometrijskih podataka.

Nekoliko postojećih alata pomažu istraživačima da izgrade i procene svoje sisteme za prepoznavanje osobe na osnovu glasa. Na osnovu rezultata prethodnih istraživanja se vidi kako funkcioniše HTK alat (Morales, 2005) i SPro3 (speech signal processing toolkit) za ekstrakciju podataka i prepoznavanje glasa (Larcher, 2013). Kako bi se izvršila interpretacija zvučnog zapisa govornika, potrebno je izvršiti izdvajanje karakteristika. Govorni signal možemo posmatrati po delovima kao kratkotrajan stacionarni signal, odnosno može se pretpostaviti da u kratkom vremenskom periodu od naprimer deset milisekundi, govor može biti shvaćen kao stacionaran proces (Khoury, 2014) (Kovačević, 2012). Prepoznavanje govornika sastoji se od detekcije govornih aktivnosti, ekstrakcije i normalizacije, dok se u pozadini vrši proces modelovanja, upisa, upoređivanja i konačne odluke. Takav proces se vidi na slici 17 (R.Klevansand, Voice Recognition, 1997) gde su prikazane glavne faze sistema za prepoznavanje govornika. Prvi korak je da se podaci koji su se prikupili pošalju putem protokola na upis, trening, odnosno evaluaciju.



Slika 17 Sistem za automatsko prepoznavanje govornika[19]

[19] Kinnunen, Tomi, and Haizhou Li. "An overview of text-independent speaker recognition: From features to supervectors." Speech communication 52.1 (2010): 12-40. (preuzeto, mart 2013)

Ovo istraživanje je izvedeno zbog toga da bi se došlo do što boljeg metoda koji bi prepoznavao osobu na osnovu glasovnog zapisa. Takođe, ovde ima i nepraktičnosti, ako se radi o manjku potrebnih informacija i zato je uvek bolje ubaciti što više podataka u bazu, da bi tačnost bila što veća. Možemo još pridoneti ako nezavisno testiramo, te na taj način garantujemo kvalitetno tržište biometrijskom tehnologijom. Velika prednost ove metode je ta što je jeftina i svima vrlo lako dostupna, mada je mnogi smatraju dopunskom biometrijskom metodom.

Dat je pregled rešenja otvorenog koda koja se mogu koristiti za implementaciju biometrijskog sistema za prepoznavanje govornika. Ovakva rešenja otvorenog koda su sve više dostupna i već postoji nekoliko softvera koji su se izdvojili svojim performansama i implementiranim algoritmima. Rešenja predstavljena u radu su Spear, MARF, Alize i HTK, koja su se izdvojila prilikom pregleda literature kao najčešće korišćene platforme otvorenog koda za prepoznavanje osoba na osnovu glasa. Prikazana je specifičnost svakog od navedenih softvera. Takođe je prikazan i model sistema za prepoznavanje govornika zasnovan na rešenjima otvorenog koda.

Najveća prednost korišćenja navedenih tehnologija prilikom implementacije sistema za prepoznavanje govornika je upravo otvoreni kod. Pre svega, ovakav kod je javno dostupan što omogućava velikom broju programera da svakodnevno unapređuju i modifikuju ove softvere. Rešenja otvorenog koda su besplatna što je samo po sebi velika prednost. S obzirom na prisutnost velikog broja ovakvih rešenja za prepoznavanje govornika moguće je realizovati kombinaciju ovakvih rešenje i dobiti kompletan, siguran i pouzdan sistem za prepoznavanje osobe na osnovu glasa.

Kinnuen i Tomi (Kinnunen T. a., 2010) su prikazali komponente sistema za automatsko prepoznavanje govornika gde se u gornjoj slici nalazi process upisa, a u donjoj process prepoznavanja govornika. Model izdvajanja karakteristika se transformiše sirov signal u vektor. U fazi upisa, model govornika je obučen da koristi za ciljanog govornika. U fazi prepoznavanja govornika, izdvajanje karakteristika (od nepoznate osobe) se upoređuju sličnosti sa podacima u bazi podataka. Konačna odluka se donosi nakon upoređivanja ovih sličnosti.

Tabela 3 predstavlja pregled upotrebe algoritama u ispitanim rešenjima otvorenog koda. Tokom istraživanja za potrebe pisanja ove doktorske disertacije došlo se do ovakvog prikaza algoritama koji su koristili ili se mogu koristiti u navedenim tehnologijama otvorenog koda.

Tabela 2 Pregled upotrebe algoritama u open source rešenjima

Program/Algoritam	FFT	MFCC	LPC	GMM	HMM (Markovljevi lanci)	Neuronske mreže
SPEAR		+		+		
MARF	+		+			+
ALIZE				+		
HTK	+	+			+	
MISTRAL	+			+		

5. VREDNOVANJE UTICAJA VARIJACIJA U INTERAKCIJI KORISNIKA SA SISTEMOM ZA PREPOZNAVANJE GOVORNIKA NA PERFORMANSE SISTEMA

Metodologija vrednovanja uticaja varijacija u interakciji korisnika sa sistemom za prepoznavanje govornika na performanse sistema, realizovanih korišćenjem gotovih i dostupnih elemenata programskih rešenja sa otvorenim programskim kodom, zasniva se na eksperimentalnim istraživanjima u kojima su merene performanse sistema u promenljivim, ali kontrolisanim, uslovima interakcije korisnika sa sistemom.

Eksperimentalna metoda je polazna tačka ovog rada. Akustička analiza govora, fonetika i logopedija, kao i druge srodne metode istraživanja predstvaljaju predmet, cilj i hipotezu samog rada. Bilo je potrebno prikupiti statističke podatke koji bi predstavili reprezentativne podatke čijom će se primenom omogućiti ispravno vrednovanje uticaja varijacija u načinu interakcije korisnika sa sistemom na performanse sistema za prepoznavanje govornika.

Sistemi za prepoznavanje govornika imaju u opštem slučaju zadatak da na osnovu ulaznih govornih podataka i odgovarajućeg skupa govornih podataka iz postojeće baze karakteristika ranije snimljenih govornika prepoznaju o kojem govorniku je reč, odnosno da zaključe da takvog govornika nema u bazi. Poseban slučaj korišćenja sistema, tipičan za rad komercijalnih sistema za prepoznavanje govornika, je rad u režimu verifikacije u kojoj govornik, pored unosa govornih podataka, tvrdi i svoj identitet, a sistem ima zadatak da prihvati ili odbaci tvrđeni identitet govornika.

Sistemi za automatsku verifikaciju govornika mogu da u procesu donošenja odluke, pored ispravnih odluka, naprave i dve vrste pogrešne odluke. Mogu da prihvate tvrđeni identitet govornika, a zapravo da bude reč o lažnom predstavljanje govornika, a mogu i da odbiju prihvatanje tvrđenog identiteta, zaključujući da je reč o slučaju lažnog predstavljanja, a zapravo nisu prepoznali ispravni identitet govornika. U zavisnosti od radne tačke sistema na ROC ili DET krivoj sistema, u opštem slučaju stope ovih grešaka će se razlikovati. Međutim, kao mera kvaliteta rada nekog sistema za automatsku verifikaciju govornika, koja omogućuje i međusobno upoređivanje kvaliteta rada više sistema, najčešće se performanse takvih sistema mere vrednošću statističkog parametra

EER, računatog po kriterijumu nalaženja radne tačke sistema u kojoj su jednake verovatnoće pogrešnog prihvatanja tvrđenog (FAR), a zapravo lažnog, identiteta govornika, odnosno pogrešnog odbijanja tvrđenog (FRR), a u realnosti ispravnog, identiteta govornika (Hofbauer & Uhl, 2016). EER je statistička mera sigurnosti zaključivanja biometrijskog sistema kojom se određuju granične vrednosti za stopu pogrešnog prihvatanja i stopu pogrešnog odbijanja tvrđenog identiteta. Za potrebe istraživanja ovog rada bilo je potrebno odabrati i pripremiti uzorke sa podacima za 30 ispitanika i sprovesti postupke testiranje radi izračunavanja EER, odnosno lokacija na ROC krivoj gde su stopa pogrešnog prihvatanja i stopa pogrešnog odbijanja jednake. Generalno, što je niža vrednost jednake stope greške, to je veća tačnost biometrijskog sistema, odnosno bolje funkcionalne performanse sistema.

U izvršenim istraživanjima merene su performanse više sistema za automatsku verifikaciju govornika realizovane programskim rešenjima otvorenog koda. Dobijene vrednosti performansi sistema u različitim radnim tačkama su korišćene radi pronalaženja radne tačke u kojoj su jednake stope greške pogrešnog prihvatanja tvrđenog identiteta i stope pogrešnog odbijanja tvrđenog identiteta. Izračunate EER vrednosti su potom upoređivane sa EER vrednostima ostalih ispitivanih sistema.

Efikasnost biometrijskog sistema ogleda se ne samo u tačnosti rada, već i u brzini donošenja odluka. Izračunavanje rezultata sličnosti između dva skupa karakteristika, ulaznog i selektovanog iz baze podataka, je osnova za odlučivanje da li oba skupa podataka pripadaju istom entitetu ili ne.

Kao što je to već ranije naglašeno, sistemi za automatsko prepoznavanje govornika na osnovu glasa pripadaju klasi multimedijalnih informacionih sistema. Za takve sisteme je karakteristično da ne rade sa potpuno interpretiranim ulaznim podacima, već su ulazni podaci tipično semistruktuirani u formi velikih binarnih objekata (BLOB) sa ograničenom semantičkom interpretacijom.

Tokom obrade posredstvom odgovarajućih programa, *interpretera,* vrši se dodatno, ali ne i potpuno, struktuiranje biometrijskih podataka. Posledica takvog pristupa obradi biometrijskih podataka je i statistička interpretacija rezultata obrade.

Upravo imajući u vidu takve karakteristike multimedijalnih informacionih sistema, različite korišćene tehnologije i načine zahvatanja biometrijskih podataka, uslove koji vladaju u okruženju u kojem se zahvatanje podataka izvodi, objektivnog i subjektivnog stanja i mogućnosti subjekta

koji daje biometrijske podatke, načina prenosa biometrijskih podataka do mesta obrade, glavna hipoteza od koje se pošlo u istraživanju je bila da tokom ponovljenih interakcija korisnika sa sistemom varijacije u ulaznim biometrijskim podacima istog subjekta mogu značajno uticati na eksploatacione performanse dostupnih programskih rešenja za prepoznavanje osoba na osnovu, pa je od posebnog interesa bilo ispitivanje osetljivosti performansi najčešće korišćenih programskih rešenja sa otvorenim kodom na varijacije ulaznih biometrijskih podataka radi pomoći u donošenju odluke prilikom realizacije jednog takvog sistema.

Tokom istraživanja trebalo je dokazati, ili odbaciti, i nekoliko posebnih hipoteza:

Pomoćna hipoteza 1: Degradacija funkcionalnih performansi sistema za prepoznavanje osobe na osnovu glasa sa rešenjem u otvorenom kodu je posledica nepodudarnosti pristupnih uređaja kao delova korisničkog interfejsa u procesima obuke i eksploatacije sistema.

Pomoćna hipoteza 2: Degradacija funkcionalnih performansi sistema za prepoznavanje osoba na osnovu glasa je posledica nepodudarnosti radnog okruženja u kojem se odvija komunikacija korisnika sa sistemom tokom obuke i tokom eksploatacije sistema.

Pomoćna hipoteza 3: Degradacija funkcionalnih performansi sistema za prepoznavanje osoba na osnovu glasa je posledica nepodudarnosti jezika na kojem korisnik komunicira sa sistemom tokom obuke i tokom eksploatacije sistema.

Pomoćna hipoteza 4: Degradacija funkcionalnih performansi sistema za prepoznavanje osoba na osnovu glasa je posledica nepodudarnosti dužine trajanja komunikacije korisnika sa sistemom tokom obuke, odnosno tokom eksploatacije sistema.

Pomoćna hipoteza 5: Degradacija funkcionalnih performansi sistema za prepoznavanje osoba je posledica semantičke nepodudarnosti zvučnih zapisa korisnika tokom obuke i tokom eksploatacije sistema.

U ovom poglavlju je opisan rezultat istraživanja na kompleksnom zadatku vrednovanja zahvaćenih biometrijskih podataka, u kome je osmišljen i primenjen algoritam za statističko vrednovanje podataka na bazi relacija između merenih veličina.

Procedura vrednovanja podataka mogla bi se metodološki podeliti u četiri koraka:

- Prvi korak: priprema potrebnih biometrijskih podataka za evaluaciju performansi sistema
- Drugi korak: generisanje izračunatih vrednosti vrednovanog podatka (izračunavanje rezultata sličnosti između dva navedena skupa karakteristika je osnova za odlučivanje da li oba skupa podataka pripadaju istom entitetu ili ne)
- Treći korak: izračunavanje EER-a;
- Četvrti korak: tumačenje izračunatih vrednosti i donošenje odluke.

U prvom koraku procedura počinje prikupljanjem i grupisanjem podataka. U ovom koraku upisuju se podaci i skladište u bazu podataka. Pretpostavlja se da su podaci potrebni za rad sistema raspoloživi i adekvatno složeni u .wav format.

U drugom koraku sistem za vrednovanje podataka se dopunjuje metodama formiranim na osnovu relacija između podataka. Metoda uključuje kako relacije između podataka, tako i procedure pretprocesiranja koje poboljšavaju karakteristike zvučnog zapisa.

Treći korak podrazumeva formiranje suda o tome da li predikcije vrednovanih podataka sadrže greške ili ne. Meri se tako da kad je u pitanju niža vrednost EER-a, to je veća tačnost sistema za prepoznavanje govornika.

U poslednjem koraku potrebno je uporediti izmerene i izračunate vrednosti, protumačiti rezultate upoređivanja i izračunati parametre po kojima se sprovodi vrednovanje podataka kao i reprezentativnu izračunatu vrednost. U ovom koraku se donosi odluka da ko je govornik.

U drugim istraživanjima nailazimo da pored EER-a može da se izvrši merenje govornika i reči za svakog govornika sa karakteristikama LPCC i MFCC kako bi se pronašle maksimalne stope identifikacije govornika i reči za obuku i testiranje klasifikatora MLP i RBF (Venkateswarlu, Raviteja, & Rajeev, 2012). Takođe, merenje može da se zasniva i na DET krivoj (Kinnunen & Li, 2010). Čini se i da duboko učenje (*eng. Deep Learning*) postaje jedno od savremenijih rešenje za verifikaciju i identifikaciju govornika (Sztahó, Szaszák, & Beke, 2019). Neki naučnici su dokazivali da je kombinacija mera definisanih preko LPCC koeficijenata i mera definisanih preko energije zaostalog signala dovela do poboljšanja u odnosu na klasičnu metodu koja uzima u obzir samo LPCC koeficijente (Faúndez-Zanuy & Rodríguez-Porcheron, 2022).

5.1.MODEL INTERAKCIJE KORISNIKA SA SISTEMOM ZA PREPOZNAVANJE GOVORNIKA

Ljudi skoro odmah prepoznaju i razlikuju glasove jedni drugih. Ali ono što je prirodno za čoveka je izazov za sistem mašinskog učenja. Kako bi rešenje za prepoznavanje govornika bilo efikasno, bilo je potrebno pažljivo izabereti model i obučiti ga na najprikladniji skup podataka sa pravim parametrima. Opet ćemo spomenuti izdvajanje karakteristika kao prvu fazu, odnosno kada sistem izdvaja osnovne glasovne karakteristike iz sirovog zvuka. Druga faza je modeliranje govornika i tada sistem kreira probabilistički model za svakog govornika. Podudaranje je treća faza kada sistem upoređuje ulazni glas sa svakim prethodno kreiranim modelom i bira model koji najbolje odgovara ulaznom glasu.

Konvertovanje govornog zvuka u format podataka koji koristi sistem je početni korak procesa prepoznavanja govornika. Prvi korak je bio snimanje ispitanika na tri uređaja u multimedijalnoj laboratoriji Fakulteta organizacionih nauka Univerziteta u Beogradu, odnosno snimanje govora eksternim mikrofonom, mikrofonom na mobilnom telefonu i integrisanom mikrofonu na laptopu, gde se zatim vršilo pretvaranje audio signala u digitalne podatke pomoću analogno-digitalnog pretvarača (u ovom istraživanju korišćen je program *Audacity*). Dalja obrada signala je uključivala procese kao što su smanjenje šuma i izdvajanje karakteristika. Deo programskog koda urađen u ovoj fazi je prikazan na slici 18. Nakon toga izvršena je normalizacija, koja je već spomenuta više puta u samom radu.

```
    from audioPaket import preprocessing
def extract_features(audio,rate):
   mfcc_feature = mfcc.mfcc(audio,rate, winlen=0.020,preemph=0.95,numcep=20,nfft=1024,ceplifter=15,highfreq=6000,
   mfcc_feature = preprocessing.scale(mfcc_feature)
   delta = calculate_delta(mfcc_feature)
   combined = np.hstack((mfcc_feature,delta))
   return combined
    return 0;
}
```

Slika 18 Deo programskog koda iz faze izdvajanje karatkeristika

Buka je neizbežno bila prisutna dok su se snimali ispitanici. Dok se snimalo mikrofonom, govorni signal je sadržao malo šuma, kao što je beli šum ili pozadinske zvukove. Prekomerna buka može izobličiti karakteristike govornih signala, degradirajući ukupan kvalitet govornog zapisa. Što više šuma sadrži audio signal, to su lošije performanse komunikacionog sistema čovek-mašina. Razvijanje svestranog algoritma za detekciju i smanjenje buke koji radi u različitim okruženjima

je izazovan zadatak, pošto su karakteristike buke nedosledne. U Alize programskom rešenju detekcija glasovne aktivnosti zasnovana je na VAD tehnici (eng. Voice Activity Detection) koja se primenjuje da bi se uklonila tišina. U eksperimentima su korišćene 32 Gausove mešavine sa dijagonalnom kovarijansom. U stvari, isprobane su mešavine od 8 do 1024. Kombinovanje GMM-a sa MFCC tehnikom ekstrakcije karakteristika pruža veliku preciznost pri izvršavanju zadataka prepoznavanja govornika (slika 19).

```
sample_rate, data = read('denoised_vad_voice.wav')

# extract 40 dimensional MFCC & delta MFCC features
features = extract_features(audio, sr)

gmm = GMM(n_components=16,max_iter=200,covariance_type='diag',n_init=1, init_params='random')
gmm.fit(features)  # gmm training
```

Slika 19 Kombinovanje GMM-a sa MFCC tehnikom

U Bob Spear i Alize programskom rešenju može se kombinovati GMM i UBM. GMM se obično primenjuje na uzorcima govora određenog govornika, izdvajajući karakteristike govora jedinstvene za tu osobu. Kada se primeni na velikom skupu uzoraka govora, GMM može naučiti opšte karakteristike govora i pretvoriti se u UBM. UBM se obično koriste u rešenjima za biometrijsku verifikaciju za predstavljanje govornih karakteristika koje su nezavisne od osobe. Kombinovani u jednom sistemu, GMM i UBM modeli mogu bolje da upravljaju govorom na koji mogu da naiđu tokom prepoznavanja govornika, poput šapata, sporog govora ili brzog govora. U Bob Spear rešenju korišćen je Kaldi. Kaldi je popularni skup alata za prepoznavanje govora za izdvajanje karakteristika (Cernak, Komaty, Mohammadi, Anjos, & Marcel). Da bi se integrisala njegova funkcionalnost sa tokovima rada zasnovanim na Python-u, koristi se paket Bob.kaldi. Primer je prikazan na slici 20.

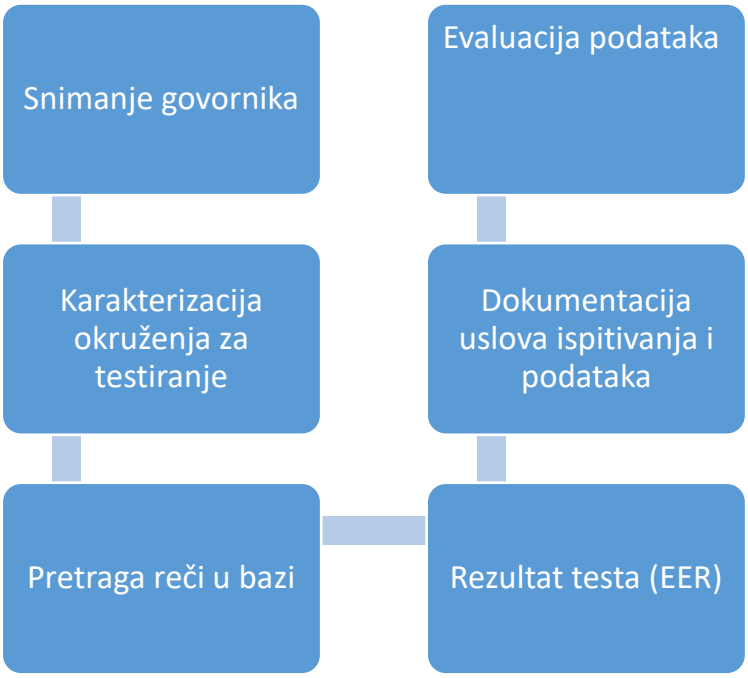
```
ubm = bob.kaldi.ubm_train(features, r'ubm_vad.h5', num_threads=4, num_gauss=2048, num_iters=100)


# training every gmm using ubm
user_model = bob.kaldi.ubm_enroll(features, ubm)


# scoring using ubm and a specified gmm
bob.kaldi.gmm_score(features, user_model, ubm)
```

Slika 20 Paket bob.kaldi

Slika 21 Sistem za prepoznavanje govornika putem programskih rešenja otvorenog koda

Način funkcionisanja sistema za prepoznavanje govornika putem programskih rešenja otvorenog koda koji su korišćeni u ovom radu je prikazan na slici 21. Na ulazni govorni signal utiče pozadinska buka. Ovaj faktor okoline može dovesti do lažnih rezultata. Uređaji za automatsko prepoznavanje govora primenjuju specifične, ali proizvoljno izabrane kriterijume odluke da bi izvršili prepoznavanje. Optimalna podešavanja ovih kriterijuma odlučivanja su izuzetno važna i kontrolisana su rečnikom, primenom sintaksičkih ograničenja, protokola za ispravljanje grešaka, karakteristikama govora i ličnih preferencija pojedinačnog korisnika.

Ono što je bitno spomenuti je da su Alize i Bob Spear podignuti na Windows operativnom sistemu, a Marf i HTK Linux operativnom sistemu. Korišćeni su i Java, C++ i Python programski jezici. Testna procedura je prikazana na slici 22.



Slika 22 Testna procedura

Ono što je potrebno izdvojiti je računanje EER. U sistemu prepoznavanja govornika jednaka stopa grešaka (EER) je mera za procenu performansi sistema. U C++ deo koda za računanje EER-a izgleda kao na slici 23, a zatim izračunavamo EER sa sledećom Python skriptom i Kaldi programom (slika 24).

```
score = []
label = []
for i in range(500):
    score.append(a[i])
    label.append(0)

for i in range(500):
    score.append(b[i])
    label.append(1)


print(len(score),len(label))
score = np.array(score)
label = np.array(label)

eer,minDcf = compute_min_cost(score, label, p_target=0.01)

print(eer,minDcf)


fpr, tpr, threshold = sklearn.metrics.roc_curve(label, score, 1)
fnr = 1 - tpr
eer_threshold = threshold[np.nanargmin(np.absolute((fnr - fpr)))]
eer_1 = fpr[np.nanargmin(np.absolute((fnr - fpr)))]
eer_2 = fnr[np.nanargmin(np.absolute((fnr - fpr)))]
# return the mean of eer from fpr and from fnr
eer = (eer_1 + eer_2) / 2
print("Jednaka stopa grešaka: ",eer,eer_threshold)
```

Slika 23 Deo koda u C++ gde se računa EER

```
eer <(python local/prepare_for_eer.py $trials exp/plda_scores/plda_scores) 2> /dev/null`
```

Slika 24 Deo koda u Python-u gde se računa EER

5.2. IZBOR ELEMENATA KORISNIČKOG INTERFEJSA I GOVORNIKA ČIJI ĆE SE UTICAJ PRATITI

Obuka ispitanika se vršila u multimedijalnoj laboratoriji na Fakultetu organizacionih nauka Univerziteta u Beogradu. Postojalo je 30 ispitanika (16 žena i 14 muškaraca), od kojih su dvadeset njih studenti Univerziteta u Beogradu, dok su ostali ispitanici zaposleni u sferama informacionih sistema i tehnologije, kao i ekonomije. Starosna grupa ispitanika je bila izmedju 18 i 35 godina. Identifikacioni podaci u bazi podataka su povezani sa pravim identitetima osoba. Ispitanici su se istovremeno snimali na tri uređaja: integrisani mikrofon na laptop računaru, profesionalni mikrofon na desktop računaru, kao i na 3G mobilnom uređaju. Snimanje se odvijalo na dva jezika, engleski i srpski. Prva grupa rečenica sadrži glasove naših slova, brojeva i naredbi za rad sa operativnim sistemom i menijem. Druga grupa su rečenice iz bankarskog domena. Treća grupa rečenica iz uobičajenog zivota (nazivi pića, hrane, doba dana, uobičajenih radnji,....). Reči i rečenice su prikazane u tabeli 3. Izjave su se snimale u kratkim (do 1 sekunde), srednjim (od 1 do 2 sekunde) i dugim zapisima (preko 2 sekunde). Pomoću tih reči/rečenica, model se obučavao. Svaki govorni signal treba da bude obeležen (snimljena kao labela) na način koji nas asocira na tekst koji je sadržan u njemu.

5.3. SCENARIJI MOGUĆIH VARIJACIJA U INTERAKCIJI KORISNIKA SA SISTEMOM

Za potrebe istraživanja ovog rada pripremljena su odgovarajuća ispitna scenarija, koja opisuju varijacije u interakciji korisnika i sistema. Na osnovu njih za svaku instancu laboratorijskog eksperimentalnog sistema za prepoznavanje govornika, formirane su različite ispitne konfiguracije elemenata sistema. Identifikacija govornika omogućava da se odredi scenario kao što je npr. „Ko govori, Ana, Milan ili Petar?“. On pripisuje govor pojedinačnim govornicima unutar grupe upisanih pojedinaca. Ispitano je kako dužina izgovorenog teksta (kratke, srednje i duge rečenice),

kao i korišćenje uređaja za snimanje (integrisani mikrofon, eksterni mikrofon i 3G mobilni telefon), te odabir jezika utiče na rezultate prepoznavanja govornika.

Neki od primera scenarija koji su ispitani:

1. scenario (korisnik dolazi u banku u kojem na šalteru stoji eksterni mikrofon i kroz razgovor računar pokušava da ga identifikuje)

Računar: „Dobar dan, kako Vam možemo pomoći?“
Govornik: „Dobar dan, potreban mi je keš kredit.“
...
Računar već u ovoj fazi počinje pretragu u bazi i pokušava da izvrši prepoznavanje govornika. Ako je korisnik identifikovan, znači da je pouzdan korisnik banke i može se dalje kreirati konverzacija u odnosu na dalje potrebe klijenta. Ako korisnik nije identifikovan, tražiće mu se konkretni lični podaci koji su potrebni banci.

2. scenario (korisnik poziva avionsku kompaniju putem mobilnog telefona kako bi rezervisao kartu)

Računar: „Dobar dan, kako Vam možemo pomoći?“
Govornik: „Dobar dan, želela bih da rezervišem let iz Beograda do Madrida u ponedeljak uveče.“
...
Računar ponovo vrši pretragu reči ili rečenica po bazi i dolazi do identifikacije osobe koja poziva i kompanija ima mogućnost da bez dodatnih parametara izvrši rezervaciju.

3. Scenario (govornik je došao iz druge države i priča engleski jezik, poziva putem laptopa u aplikaciji Skajp svog poslovnog partnera da se doda novi korisnik koji može koristiti određene opcije na svom računaru)

Računar: „Good afternoon, how can I help you?“
Govornik: „Hello, can you add user Petar Petrovic for business operations?“
...
Računar vrši prepoznavanje govornika, proverava ko je osoba koja govori i da li se nalazi u bazi, i prati da li ima ovlašćenja da doda novog korisnika ili ne.

6. STUDIJA SLUČAJA: ISPITIVANJE UTICAJA VARIJACIJA U INTERAKCIJI KORISNIKA SA SISTEMOM NA PERFORMANSE SISTEMA ZA PREPOZNAVANJE GOVORNIKA

Osnova rada sistem za prepoznavanje govornika su biometrijski podaci o glasu kao zvučnom podatku govornika. Da bi se zvuk registrovao, mora postojati signal koji ga prenosi. Nakon analogno-digitalne konverzije govornog signala, sistem raspolaže digitalnim oblikom glasa koji može da se procesira ili zapiše na memorijskom mediju. Kao za većinu multimedijalnih entiteta, za zapisivanje zvučnih podataka potrebno je znatno više memorije nego što je to slučaj kod transakcione obrade podataka, pa se često primenjuje kompresija takvih podataka. Međutim, potrebno je obratiti pažnju na kompresiju zvučnog zapisa, jer se redukovanjem memorije smanjuje i vernost originalnog zvučnog zapisa.

Formati zvučnih zapisa se dele na (Kuršumlija, 2012):

- nekompresovane (MIDI, WAV, AIFF)
- kompresovane bez gubitaka (WMA, FLAC)
- kompresovane sa gubicima (MP3, Ogg Vorbis, AAC)

Format koji je u radu korišćen je WAV. WAV matični format razvijen je za korišćenje u Windows okruženju i podržan od strane većine web pregledača. Datoteke u ovom formatu imaju nastavak .wav.

Bez obzira da li je u pitanju trening ili test faza, prepoznavanje glasa izvodi se u nekoliko opštih koraka:

1. Snimanje glasa
2. Pretprocesiranje i konvertovanje u digitalni oblik
3. Izdvajanje karakteristika
4. Klasifikacija i
5. Donošenje odluke

Među ovim koracima, izdvajaju se tri, koji se mogu realizovati na više načina, implementiranjem nekoliko algoritama. To su pretprocesiranje, izdvajanje karakteristika i klasifikacija. U zavisnosti

od kombinacije ovih koraka, pouzdanost i preciznost sistema za prepoznavanje glasa može biti manja ili veća. Kako težimo da efikasnost sistema bude što veća, odabir ove kombinacije je od krucijalne važnosti.

6.1. RADNO OKRUŽENJE ZA FORMIRANJE BAZE ZVUČNIH ZAPISA

Podaci koji su se prikupili za potrebe ovog rada su snimljeni na srpskom i na engleskom jeziku. Takođe, snimljeni su na tri uređaja: Integrisani mikrofon na laptop računaru (prenosni računar), profesionalni mikrofon koji je priključen na desktop računar i 3G telefon. Detalji su navedeni u Tabeli 4. Govorni podaci su se snimali istovremeno na sva tri uređaja u multimedijalnoj laboratoriji Fakulteta organizacionih nauka.

Tabela 3 Specifikacije uređaja za snimanje

Uređaj	Konfiguracije	Format
Desktop računar sa eksternim mikrofonom	Model računara: Fujitsu, Intel(R) Core(TM) i5-2320 CPU	WAV
	Model mikrofona: Hyundai CJC-M30	
	Audio: stereo 24 bit Raspon frekvencije: 20-20000 Hz	
3G telefon	Model: Samsung S4 (GT-I9505)	WAV
	Audio: stereo 24 bit Raspon frekvencije: +0.03, -0.08	
Laptop računar sa integrisanim mikrofonom	Model: HP ProBook 650 G1, Intel(R) Core(TM) i3-4000M CPU	WAV
	Audio: stereo 24 bit Raspon frekvencije: 48000 Hz	

6.2. FORMIRANJE TEKSTUALNIH LISTA ZA BAZU ZVUČNIH ZAPISA

Osnovni princip rada alata koji su se koristili u ovom radu može se objasniti u par koraka – postoje dva osnovna procesa koji se odvijaju sledno: trening i prepoznavanje. Pre nego što pokrenemo oba procesa, bitno je dobro pripremiti sve podatke potrebne za prepoznavanje. Koraci su sledeće:

- Priprema i organizovanje podataka – vrši se snimanje uzoraka za trening i test, te se kreiraju transkripcije. Zatim se definiše vokabular koji se kasnije upotrebljava za prepoznavanje osobe. Potom se izrađuje rečnik izrađuje se rečnik transkripcija na temelju vokabulara materijala za trening i test, te se rekonstruiše govorni signal (npr. wav u mfc);
- Treniranje – definišu se modeli, odnosno vrsta modela koja zavisi od načina treniranja (metoda izolovanih reči ili metoda slednih reči)
- Prepoznavanje – posle izgradnje modela i upotrebe gramatike (mogući redosled reči) sprovodi se transkripcija neidentifikovanih reči/rečenica i onda dolazi do analize rezultata.

6.3. IZBOR I OBUKA ISPITANIKA

Obuka ispitanika se vršila u multimedijalnoj laboratoriji na Fakultetu organizacionih nauka Univerziteta u Beogradu. Starosna grupa ispitanika je bila izmedju 18 i 35 godina. Ispitanici su se istovremeno snimali na tri uređaja: integrisani mikrofon na laptop računaru, profesionalni mikrofon na desktop računaru, kao i na 3G mobilnom uređaju. Govornici tečno pričaju srpski i engleski jezik, te se snimanje odvijalo na oba jezika. Postojalo je 30 ispitanika (16 žena i 14 muškaraca). Postojale su tri grupe reči/rečenica za snimanje.

Ispitanici su pozivani i svaki je individualno snimljen. Puštane su im prezentacije na oba jezika, te su iščitavali prikazane reči, odnosno rečenice. Istovremeno su bila uključena sva tri uređaja. Nakon toga podaci su se ubacivali u bazu podataka.

6.4. SNIMANJE ZVUČNIH ZAPISA

Za funkcionisanje sistema za prepoznavanje glasa, dovoljno je postojanje računara sa adekvatnim softverom i običan mikrofon, a u ovom slučaju i spomenuti mobilni uređaj. Da bismo dobili bolje rezultate prepoznavanja, preporučuje se korišćenje istog modela mikrofona i u fazi upisa i u fazi testiranja. Format snimka je 24-bitne dubine i nivoa uzorkovanja je minimum 8000 Hz. Snimljeni

glas se zatim podvrgava koraku pretprocesiranja, gde se uklanja šum, normalizuje analogni signal, a nakon toga se odabranim algoritmom (obično LPC ili Filterbank) konvertuje u parametarski oblik. Najbitnija parametarska reprezentacija govora je kratkoročni spektralni okvir (eng. *short time spectral envelope*).

Za potrebe snimanja, kreirana su tekstualna uputstva na srpskom i engleskom jeziku. Izjave se snimaju u kratkim (do 1 sekunde), srednjim (od 1 do 2 sekunde) i dugim zapisima (preko 2 sekunde). Tekstualna uputstva za izgovor se dele u tri kategorije:

- Prva grupa rečenica sadrži glasove naših slova, brojeva i naredbi za rad sa operativnim sistemom i menijem.
- Druga grupa su rečenice iz bankarskog domena.
- Treća grupa rečenica iz uobičajenog zivota (nazivi pića, hrane, doba dana, uobičajenih radnji,....)

Tabela 4 Skupovi reči i rečenica pri snimanju govornika

	Izvorno:	***Prevod:***
Grupa 1		
Skup 1		
1	jedan	one
2	dva	two
3	tri	three
4	četiri	four
5	pet	five
6	šest	six
7	sedam	seven
8	osam	eight
9	devet	nine
10	nula	zero
11	a	
12	c	
13	f	
14	i	
15	c	

16	plus	plus
17	minus	minus
18	saberi	add
19	oduzmi	subtract
20	pomnoži	multiply
21	podeli	divide
22	otvori	open
23	zatvori	close
24	odustani	cancel
25	snimi	save
26	obriši	delete
27	kopiraju	copy
28	premesti	move
29	klikni	click
30	preimenuj	rename
Skup 2		
1	označi sve	select all
2	kopiraj tekst	copy text
3	unesi šifru	enter password
4	promeni šifru	change password
5	korisničko ime	username
6	internet brzina	internet speed
7	količina podataka	data size
8	kopiraj datoteku	copy file
9	nalepi tekst	paste text
10	nalepi datoteku	paste file
11	promeni naziv	change name
12	promeniti pozadinu	change background
13	ažuriraj sistem	system update
14	koristi prečicu	use shortcut
15	definiši prečicu	define shortcut
16	maksimizuj prozor	maximize window
17	minimizuj prozor	minimize window
18	tačka povratka	restore point
19	inicijalizacija objekta	object initialization
20	snimi kao	save as

21	dodaj korisnika	add user
22	obriši korisnika	delete user
23	izmeni korisnika	edit user
24	novi jezičak	new tab
25	obriši istoriju	clear history
26	nedavno zatvoreno	recently closed
27	menadžer zadataka	task manager
28	novi prozor	new window
29	ubaci red	insert row
30	ubaci kolonu	insert column
Skup 3		
1	pokreni kao administrator	run as administrator
2	novi privatni prozor	new incognito window
3	uvezi markere i podešavanja	import bookmarks and settings
4	da li ste sigurni da želite da obrišete dokument	are you sure you want to delete the document
5	ukucajte ovde da biste tražili	type here to search
6	promenite podešavanja za fascikle i pretragu	change folder and search options
7	vindouz operativni sistem	the windows operating system
8	trenutni korisnički grafički interfejs	the current GUI interface
9	tip korisnički prijateljskih operativnog sistema	type of "user friendly" operating system
10	učitavanje operativnog sistema	loading the operating system
11	markiraj otvorene stranice	bookmark open pages
12	prikaži liniju za markerima	show bookmarks bar
13	korsiti polje za potvrdu za odabir elemenata	use checkboxes to select items
14	sakrij konflikte nakon spajanja datoteka	hide folder merge conflicts
15	sakrij ekstenzije za poznate tipove fajlova	hide extension for known file types
16	prikaži ikonu fajla na sličicama	display file icon on thumbnails
17	prikaži punu putanju u naslovu	display the full path in the title bar
18	pošalji hitni e-mail sa posla	send an urgent email from my work
19	budi siguran da me cortana čuje	make sure cortana can hear me
20	vaš uređaj je ažuriran	your device is up to date

21	tražite informacije o poslednjim ažuriranjima	looking for info on the latest updates
22	nadograditi svoju verziju vindouza	upgrade your edition of windows
23	pokrenite sa uređaja ili diska	startup from a device or disk
24	naučite kako da počnete sa novom instalacijom	learn how to start fresh with clean installation
25	da li prihvatate uslove	do you accept terms and conditions
26	sve izmene su sačuvane	all changes saved in drive
27	postavite limit za snimanje na pet megabajta	set download limit to five mb
28	postavite limit za slanje na tri megabajta	set upload limit to three mb
29	izmenite predefinisanu lokaciju za snimanje datoteka	change download file default location
30	osigurajte se da je uređaj priključen putem usb-a	make sure your device is visible to usb connections
Grupa 2		
Skup 1		
1	pin	pin
2	kod	at
3	račun	bill
4	broj	number
5	šalter	window
6	ističe	highlights
7	kredit	credit
8	bankomat	ATM
9	banka	bank
10	kamata	interest
11	štednja	savings
12	klijent	client
13	red	red
14	ekspozitura	branches
15	keš	cache
16	žirant	endorser
17	stambeni	residential
18	grejs	grace
19	partner	partner
20	euribor	euribor

21	ponuda	offer
22	sef	strongbox
23	terminal	terminal
24	efektiva	effective
25	privilegije	privileges
26	dokumentacija	documentation
27	broker	broker
28	kupoprodaja	buying and selling
29	isplata	payment
30	investicija	investment
Skup 2		
1	Tekući račun	Current account
2	Dokumentacija	documentation
3	Cena kredita	Price of credit
4	Mesečna provizija	Monthly fee
5	Keš kredit	Cash loan
6	Kamatna stopa	Interest rate
7	Dozvoljeni minus	overdraft
8	Prazan formular	empty form
9	Novac u inostranstvo	Money abroad
10	Devizni račun	Foreign currency account
11	Stambeni kredit	Housing loan
12	Oročena Štednja	Term Savings
13	e-banking aplikacija	e-banking aplikacija
14	Kursna lista	Exchange rate
15	Fizičko lice	Physical face
16	Trajni nalog	Standing order
17	Nova kartica	The new card
18	Gubitak ili krađa kartice	Lost or stolen cards
19	Limit na bankomatima	Limit on ATMs
20	Bezbednost kartice	safety cards
21	isplata efektive	cash payments
22	uplata efektive	cash payments
23	prenos neto zarada	transfer of net earnings
24	prihvatanje platnih kartica	accepting payment cards
25	kredit za refinansiranje	loan for refinancing

26	brokersko dilerske usluge	broker-dealer services
27	digitalni bankarski servis	digital banking service
28	servisne informacije	service information
29	aktuelne kampanje	current campaign
30	finansijski izveštaj	financial report
Skup 3		
1	Ko može otvoriti tekući račun u Banci?	Who can open an account at the Bank?
2	Šta je potrebno od dokumentacije?	What documents are needed?
3	Koja je cena otvaranja tekućeg računa ?	What is the cost of opening a current account?
4	Kolika je mesečna provizija za vođenje tekućeg računa?	What is the monthly fee for keeping the current account?
5	Da li nudite dinarske keš kredite?	Do you offer dinar cash loans?
6	Koja je kamatna stopa za kredite ročnosti do 12 meseci?	What is the interest rate for loans with maturity up to 12 months?
7	Šta mi je od dokumentacije potrebno za uslugu dozvoljenog minusa?	What is the required documentation for overdraft
8	Mogu li dobiti prazan formular za uslugu dozvoljenog minusa?	Can I get a blank form for the service of overdraft?
9	Na koji način mogu poslati novac u inostranstvo?	How can I send money abroad?
10	Ko može da uplati novac na moj devizni račun?	Who can pay the money to my foreign account?
11	Koji su uslovi za podnošenje zahteva za stambeni kredit?	What are the requirements for applying for a mortgage loan?
12	Koji dokumenti su mi potrebni za štednju?	What documents do I need to save money?
13	Da li preko vase e-banking aplikacije mogu aplicirati za kredit?	Can I apply for a loan by using e-banking application?
14	Da li vasa e-banking aplikacija nudi kursne liste?	Does your e-banking application offers exchange rate list?
15	Kolika je mesečna provizija za održavanje računa za fizička lica?	What is the monthly fee for account maintenance for individuals?
16	Interesuje me da li nudite uslugu trajnog naloga i pod kojim uslovima?	I wonder if you offer a service standing orders and under what conditions?
17	Kartica mi ističe ovog meseca. Da li moram ponovo zahtevati novu karticu?	My card expires this month. Do I need to re-request a new card?
18	Šta da radim u slučaju gubitka ili krađe kartice?	What should I do in case of loss or theft?

19	Koji je limit za podizanje gotovine na bankomatima?	What is the limit for cash withdrawals at ATMs?
20	Da li je bezbedno koristiti karticu na Internetu?	Is it safe to use the card on the Internet?
21	Plaćajte mobilnim telefonom umesto karticom	Pay with mobile phone instead of card
22	Bezbedna kupovina preko interneta uz Verified by Visa tehnologiju	Safe Buying on line with Verified by Visa technology
23	Brzo i jednostavno plaćanje u zemlji i inostranstvu na prodajnim mestima opremljenim beskontaktnom tehnologijom	Quick and easy payment in the country and abroad at points of sale equipped with contactless technology
24	U našoj banci možete unovčiti obveznice stare devizne štednje	In our bank, you can cash in foreign currency savings bonds
25	Nudimo usluge zakupa sefova na više lokacija u našim ekspoziturama	We offer services of renting safe deposit boxes at several locations in our branches
26	Elektronski samoposlužujući sefovi predstavljaju poslednju reč tehnologije	Electronic self safes represent the latest technology
27	Više ne morate da dolazite u banku da biste platili mesečni račun	You no longer have to come to the bank to pay a monthly bill
28	Izvršite bankarske transakcije putem telefonskog razgovora sa agentom kontakt-centra Banke.	Perform banking transactions via telephone conversation with Agent contact-center Bank.
29	Svakog meseca s vašeg tekućeg računa plaćamo vaše mesečne račune.	Each month we use your checking account to pay your monthly bills.
30	Menjački poslovi obuhvataju otkup i prodaju efektivnog stranog novca bez provizije	Exchange operations include the purchase and sale of foreign cash without commission
Grupa 3		
Skup 1		
1	putovanje	travel
2	ručak	lunch
3	dezert	dessert
4	kuća	house
5	spavanje	sleep
6	vikend	weekend
7	posao	work
8	praznik	holiday
9	vreme	time
10	restoran	restaurant

11	pasoš	passport
12	glas	voice
13	prodavnica	shop
14	cipele	shoes
15	prijatelj	friend
16	poznanik	acquaintance
17	doručak	breakfast
18	večera	evening
19	užina	brunch
20	voće	fruit
21	sport	sport
22	trčanje	jogging
23	poseta	visit
24	komšija	neighbor
25	terasa	feel
26	kafa	coffee
27	čaj	tea
28	vino	vine
29	kupanje	swimming
30	gledati	watch
Skup 2		
1	moj glas je moja šifra	My voice is my code
2	razmatranje ponude	considering offer
3	poseta porodici	visiting family
4	porodični ručak	family lunch
5	popodnevni odmor	afternoon rest
6	jutarnja šetnja	morning walk
7	poseta muzeju	visit the museum
8	slušanje muzike	listening to music
9	odlazak u bioskop	going to the cinema
10	poseta pozorištu	visiting the theater
11	odmor u parku	resting in the Park
12	čitanje knjiga	reading books
13	vožnja bicikla	cycling
14	vožnja rolera	skating
15	piknik na reci	picnic on the river

16	pospremanje kuće	cleaning the house
17	pranje automobila	car wash
18	izlazak sa društvom	out with friends
19	poslovni ručak	business lunch
20	vožnja tramvajem	tram ride
21	odlazak na pijacu	going to the market
22	neradni dan	non-working day
23	godišnji odmor	vacation
24	zaliti cveće	water the flowers
25	gledanje utakmice	watching the game
26	pranje veša	laundry
27	popij čaj	drink tea
28	idemo kući	Let's go home
29	poruči burger	order burger
30	parkiraj automobil	park the car
Skup 3		
1	Spremimo se pre nego što stignu gosti	Get ready before the guests arrive
2	Zatvori prozor da ne duva promaja	Close the window, there is a draft.
3	Moje radno vreme počinje u 8 ujutro	My working time starts at 8 AM
4	Moje radno vreme završava u 4 popodne	My working time ends at 4 pm
5	Treniraj barem tri puta nedeljno	Workout at least three times a week
6	Pogledaj koliko je sati, kasnimo.	Look at the time, we're late.
7	Za zdraviji život jedi više voća	For a healthier life eat more fruit
8	Kako vam se dopada pozorišna predstava od prošle nedelje	How do you like the theater performance from last week
9	nemoj da preskočiš odlazak kod frizera	do not skip going to the hairdresser
10	jednom godišnje obavezno idem kod lekara na kontrolu	i visit my doctor annually in order to perform regular test.
11	svake nedelje odlazim na pijacu	every sunday I go to the market
12	Volim da pripremam ručak od svežih namirnica	I like cooking lunch of fresh foods
13	volim da popijem čašu vina u sumrak	I like to drink a glass of wine at sunset
14	koja je tvoja omiljena knjiga	What's your favorite book
15	koliko dnevno popijemo litara vode	how much liters of water do we drink per day
16	naše komšije su se doselile prije par dana	Our neighbors have moved in a few days ago

17	treba da platimo račune do desetog u mesecu	we need to pay the bills by the tenth of the month
18	nismo imali priliku da se upoznamo ranije	We have not had the opportunity to meet earlier
19	na graničnim prelazima je gužva celi dan	border crossings is jammed all day long
20	bolje je ranije početi sa planiranjem godišnjeg odmora	It is better to start earlier with vacation plans
21	posedujem mesečnu kartu za gradski prevoz	I own a monthly ticket for public transport
22	više se isplati kupiti povratnu kartu	makes more sense to buy a return ticket
23	u našem hotelu doručak služe do 9 sati	in our hotel breakfast is served until 9 am
24	koja je poenta filma koji smo gledali sinoć	what is the point of the film that we watched last night
25	proleće je najlepše godišnje doba	Spring is the best time of the year
26	moramo da platimo ratu kredita	we have to pay the loan installment
27	želeo bih da rezervišem let	I would like to book a flight
28	iz Beograda do Madrida preko Ciriha	from Belgrade to Madrid via Zurich
29	odlučio sam da puštam kosu	I decided to let my hair grow
30	kupi mleko, hleb, jaja i slaninu	buy milk, bread, eggs and bacon

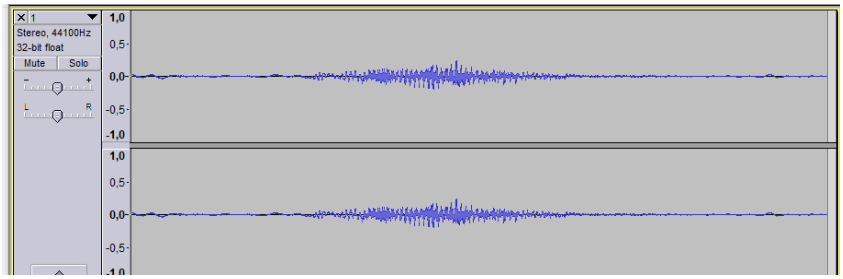
6.5. PRETPROCESIRANJE ZVUČNIH ZAPISA

Procesiranje dolazi posle snimanja, a ono je zaduženo za pripremu i prebacivanje analognog zvučnog signala u digitalni i parametarski oblik. U ovoj fazi se obrađuje signal, gde se snimci snimljeni u različitim uslovima, normalizuju i svode na sličan oblik. Pretprocesiranje se isto tako sastoji od nekoliko faza:

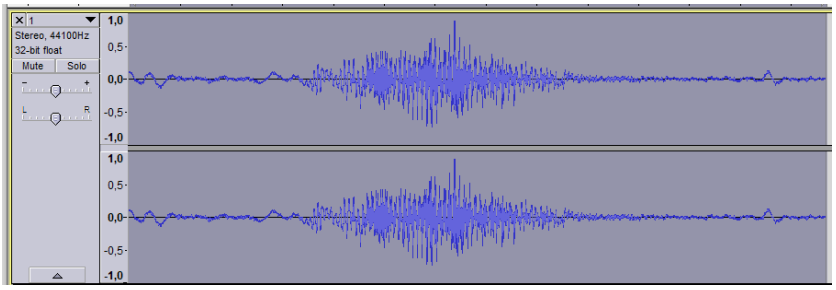
- *Normalizacija* kojoj se podvrgavamo je da bi se uklonio problem sa snimljenim uzorcima koji su različitih jačina. Prvo se pronalazi maksimalna amplituda u uzorku, a onda se skalira uzorak tako što se svaka tačka podeli sa ovom maksimalnom vrednošću. Na taj način dobija se normalizovani uzorak koji je spreman za dalju obradu. Slike 18-19 predstavljaju prikaz audio snimka pre i posle normalizacije.

  Postoje novija istraživanja normalizacije gde se spominju novi algoritmi normalizacije

poput PCMN (eng. *Parametric Cepstral Mean Normalization)* kako bi se omogućilo povećanje robustnosti prepoznavanja govora ili govornika u različitim akustičnim uslovima (Kalinli, Bhattacharya, & Weng, 2019).

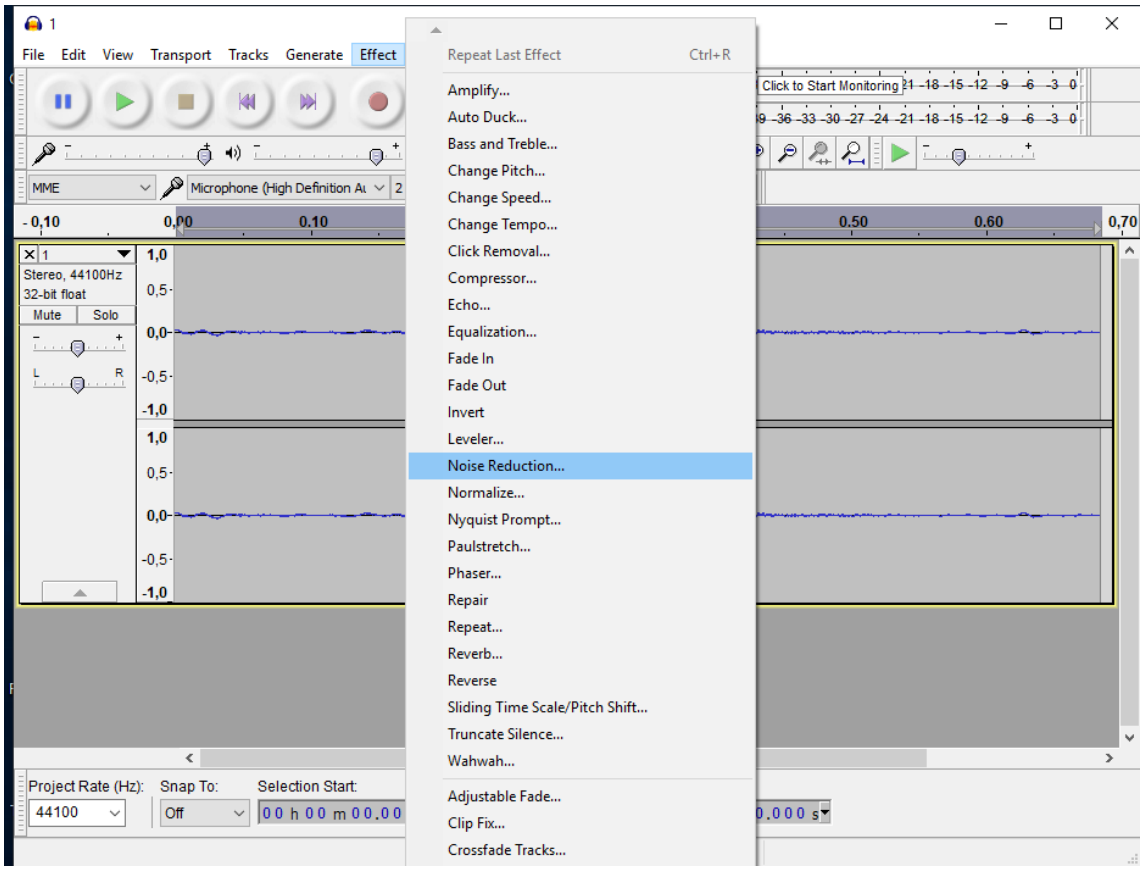


Slika 25 Audio snimak pre normalizacije

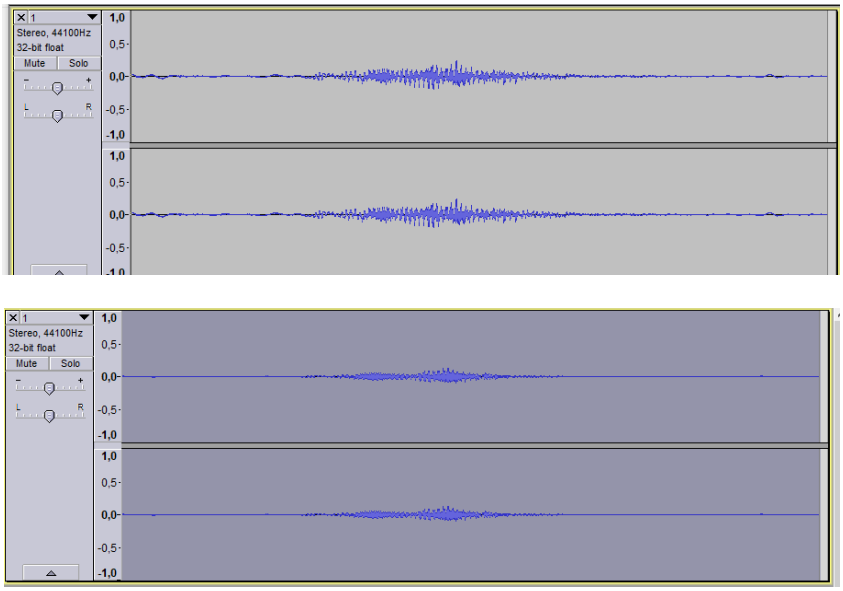


Slika 26 Audio snimak posle normalizacije

- *Uklanjanje šuma* je proces koji nam služi da očistimo i uklonimo zvukove koji nam smetaju, koji su nastali zbog loših uslova snimanja ili interferencija.

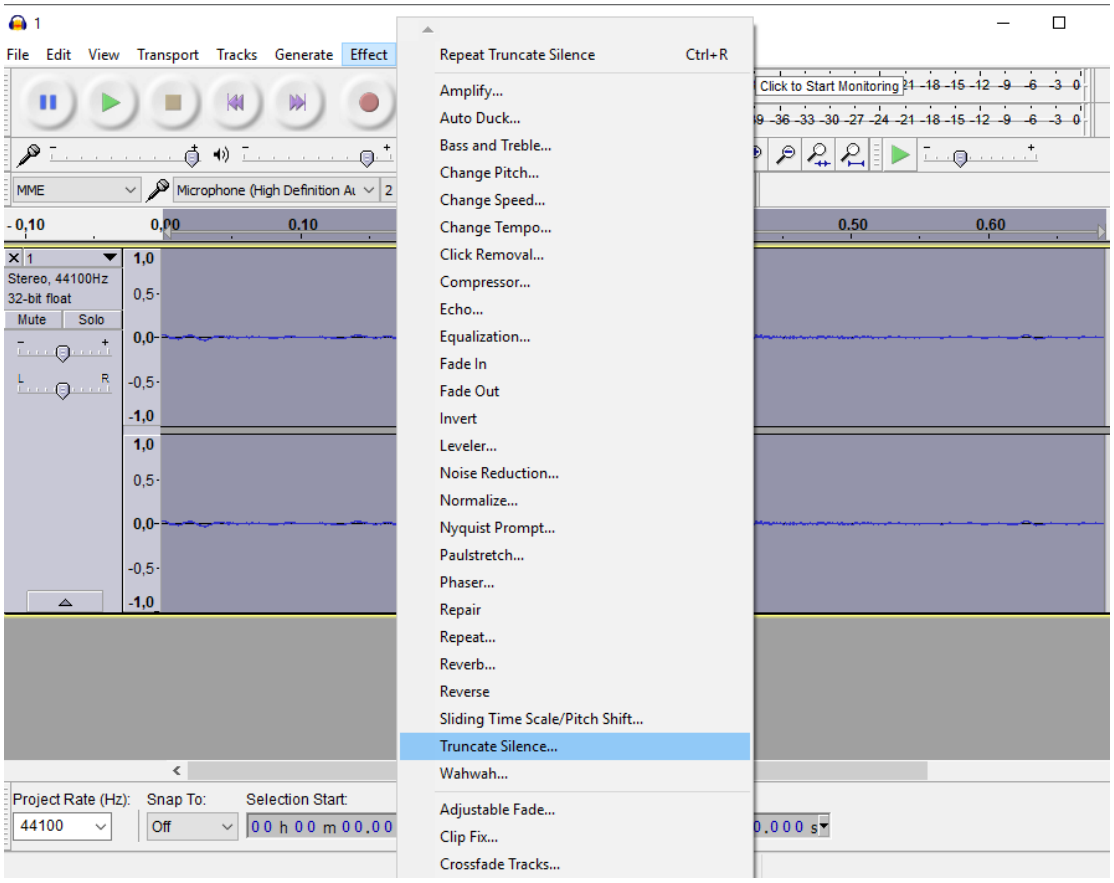


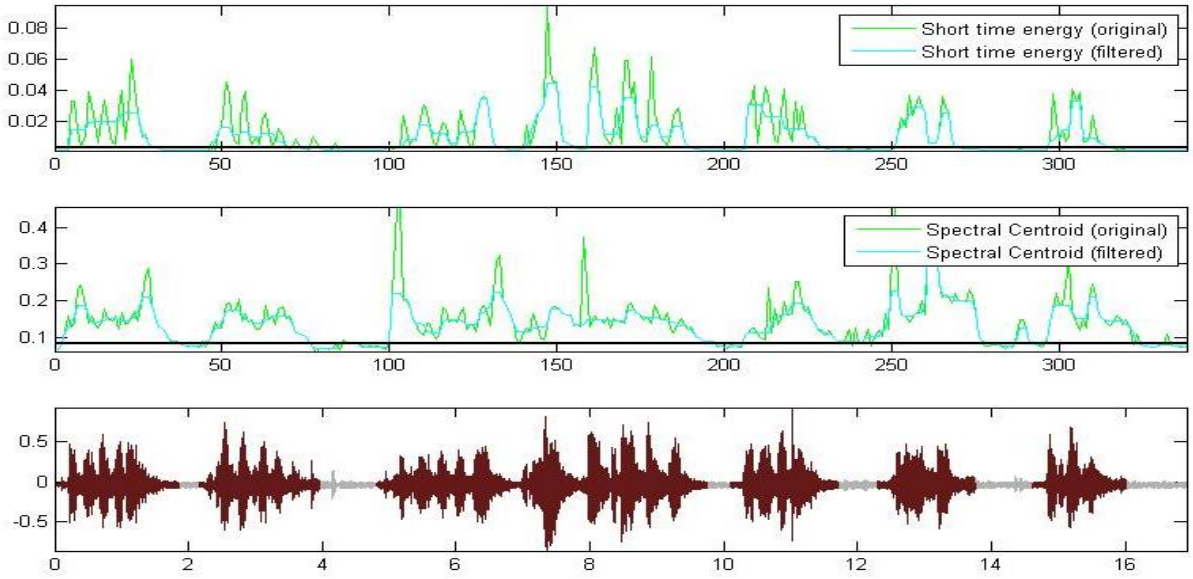
Slika 27 Funkcija uklanjanja šuma



Slika 28 Audio snimak sa postojanjem šuma i nakon njegovog uklanjanja

- *Uklanjanje tišine* se odnosi na to da izdvojimo samo delove zvuka koji su nam od značaja, tj. one koji ne predstavljaju glas, odnosno govor. Ovim načinom smanjujemo veličinu uzorka i modifikujemo ga tako da poseduje manje sličnosti sa drugim uzorcima. Ovaj postupak se vrši u vremenskom domenu i najbolje ga je sprovesti nakon postupka normalizacije kako bi njegovim korišćenjem poboljšali celokupne performanse sistema za prepoznavanje glasa.

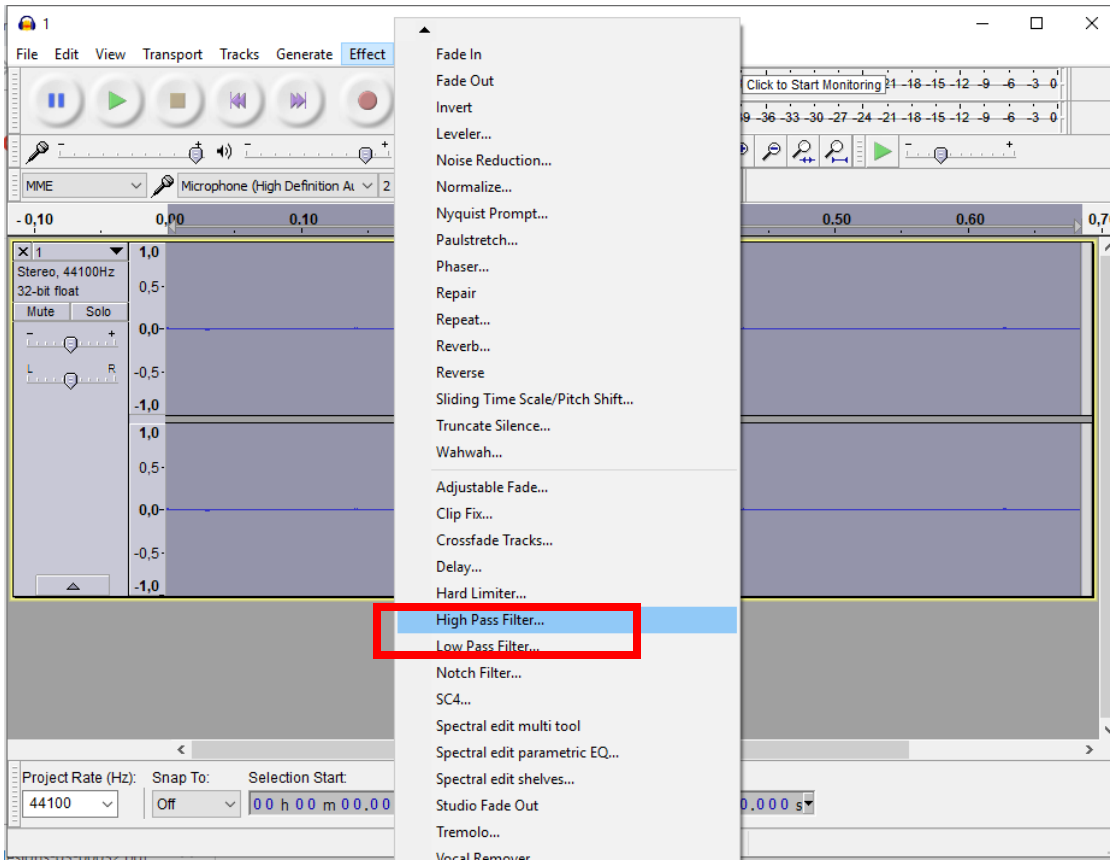




Slika 29 Metod uklanjanja tišine[20]

- *Filteri* nam služe kako bi se što bolje izmerile pojedine frekvencije koje nas interesuju. Za prepoznavanje glasa su od posebnog značaja podsticaj visokih frekvencija (eng. high frequency boost) i low-pass filter koji prolazi kroz elektronski niske frekvencije i smanjuju amplitudu frekvencije višim od prekida frekvencije.

[20] http://www.mathworks.com/matlabcentral/fileexchange/authors/30223



Slika 30 Funkcija filtera (High Pass Filter i Low Pass filter)

- *Endpointing* (krajnje tačke) predstavlja lokalne minimume i maksimume u promenama amplitude. To je process pronalaženja ovih krajnjih tačaka u nekom zvučnom zapisu, kako bi se odredilo postojanje govora u pozadinskom šumu. Korišćenjem ove tehnike mogu se izdvojiti početak i kraj reči, pa se ona često koristi i u procesu prepoznavanja govora. Da bi funkcionisao kako treba, pri implementaciji ovog algoritma treba obratiti pažnju na nekoliko specijalnih situacija kao što su: reči koje počinju ili se završavaju sa fonemama slabe energije; reči koje se završavaju sa nazalnim slovom; govornici koji se pri izgovoru reči utišavaju.

6.6. ORGANIZACIJA I SKLADIŠTENJE PRETPROCESIRANIH ZVUČNIH ZAPISA

Kao što je već spomenuto, dok se vrši snimanje govornih zapisa, često govor sadrži i šum, odnosno pozadinsku buku. Prisutnost šuma može da umanji tačnost klasifikacije, te je potrebno izvršiti pretprocesiranje.

Klasifikacija predstavlja proces donošenja odluke o tome, ko je osoba koja je tvorac glasovnog signala, na osnovu prethodno sačuvanih i naučenih informacija. Ovaj proces se obično deli u dva koraka:

- Uparivanje
- Modelovanje

Modelovanje je, u suštini, proces upisivanja osobe u sistem kreiranjem modela njenog glasa na osnovu karakteristika izdvojenih iz uzoraka njenog glasa u prethodnom koraku. Uparivanje je proces izračunavanja ocene uparivanja, koja predstavlja meru sličnosti karakteristika izdvojenih iz uzorka nepoznate osobe sa modelom nekog od govornika iz baze. Postoje dva pristupa klasifikaciji u oblasti prepoznavanja govora:

- Uparivanje šablona
- Stohastičko uparivanje

Euklidsko odstupanje predstavlja rastojanje između dve tačke i definisano je Pitagorinom teoremom. U slučaju vektora izdvojenih karakteristika, koristi se za pronalaženje razlike između dva takva vektora. Onaj vektor koji najmanje odstupa od trening vektora, predstavlja najbolji pogodak u identifikaciji.

Diff odstupanje je izmišljeno od strane jednog od začetnika MARF projekta. Glavna stvar kod ovog klasifikatora je da izračuna koliko se jedan ulazni vektor razlikuje od drugog u pogledu odgovarajućih elemenata.

Chebyshevljevo odstupanje je maksimalno odstupanje između dva vektora po bilo kojoj koordinatnoj dimenziji. Ako imamo npr. 2 tačke x(0,5) i y(7, 10), Chebyshevljevo odstupanje max(| 7 – 0 | , | 10 – 5 |), tj. 7. Ovaj algoritam se često naziva „šahovska razdaljina“ jer je u igri šah minimalan broj poteza potrebnih da kralj dođe od jednog do drugog polja na tabli jednak Chebyshevljevom odstupanju centara kvadrata.

Skriveni model Markova predstavlja statistički Markovljev model u kojem se modelira sistem za koji se pretpostavlja da je Markovljev proces sa neposmatranim tj. skrivenim stanjima. Za razliku od standardnih Markovljevih modela, gde je stanje direktno vidljivo posmatraču, kod skrivenih modela stanja nisu direktno vidljiva, dok je izlaz, koji je zavistan od ovih stanja, vidljiv. HM modeli su posebno upotrebljivi u prepoznavanju vremenskih šablona kao što je slučaj u tehnologiji prepoznavanja glasa. Uopštena arhitektura skrivenih Markovljevih modela je takva da postoje proizvoljne promenljive koje mogu uzeti bilo koju promenljivu iz niza vrednosti. Proizvoljna promenljiva x(t) predstavlja skriveno stanje u vremenu t, dok proizvoljna primenljiva y(t) predstavlja jedno posmatranje u vremenu t. Skriveno stanje x(t) zavisi samo od skrivenog stanja x(t-1), a ne i od prethodnih stanja. Takođe, posmatranje y(t) zavisi samo od skrivenog stanja x(t). Ovo zapažanje predstavlja svojstvo Markova. U HM modelu prostor stanja skrivenih promenljivih je diskretan, dok posmatranja mogu biti bilo diskretna bilo kontinualna. Postoje 2 tipa parametara u HM modelu, verovatnoće prelaza i emisione verovatnoće. U tehnologiji prepoznavanja glasa HM modeli imaju veliku primenu. Koriste se kako bi se glas prikazao kao X n-dimenzionih vektora realnih vrednosti, gde broj ovih vektora zavisi od parametra Y koji ukazuje na to nakon koliko vremena (obično 10ms) će se izračunati novi vektor. Ovi vektori se sastoje od ranije pomenutih cepstralnih koeficijenata.

Kvantizacija vektora je jedan od najjednostavnijih tekstualno nezavisnih modela govornika. Njegova primena je počela 1980-ih. S obzirom na lakoću implementiranja i nezahtevnost u pogledu iskorišćenja resursa često se koristi za brze računarske proračune, ali takođe daje zadovoljavajuću preciznost u prepoznavanju glasa kada se koristi u kombinaciji sa pozadinskim modelima. U procesu prepoznavanja glasa kvantizacija vektora predstavlja proces prtvaranja velikog seta vektora karakteristika u manji set mernih vektora koji predstavljaju centre distribucije. Korišćenjem ove metode, karakteristike iz trening podataka se grupišu i formiraju šifrarnik za svakog govornika. U procesu prepoznavanja, upoređivanjem test podataka sa ovim šifrarnikom dobija se odstupanje, na osnovu koga se kasnije donosi odluka.

Gausovi mešoviti modeli su jedna od najboljih statističkih metoda za grupisanje podataka. Može se smatrati da su GMM proširenje prethodno opisanog VQ modela, kod koga se grupe (klasteri) preklapaju. U tehnologiji prepoznavanja glasa, svaki govornik se može reprezentovati Gausovim mešovitim modelom.

Parametri GMM modela se mogu odrediti metodom procene maksimalne verovatnoće (eng. Maximum Likelihood). Glavni cilj ML procene je pronalaženje optimalnih parametara modela, koji će maksimizirati verovatnoću Gausovog mešovitog modela. S obzirom da direktna maksimizacija korišćenjem ML procene nije moguća, koristi se specijalan slučaj ove procene koji se zove maksimizacija očekivanja (eng. Expectation Maximization). GMM verovatnoća niza od T trening vektora $X = \{\overline{x_1}, \dots, \overline{x_T}\} X = \{\overline{x_1}, \dots, \overline{x_T}\}$ može se zapisati kao:

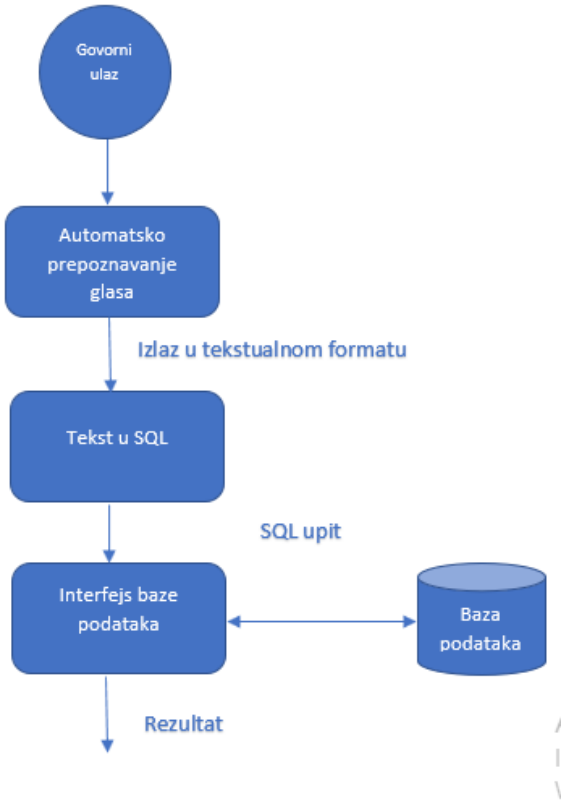
$$p(X \mid \lambda) = \prod_{t=1}^{T} p(\overline{x_T} \mid \lambda)$$

EM algoritam uzima početni model $\lambda\lambda$ i teži da pronađe novi model $\bar{\lambda}\bar{\lambda}$ takav da je

$p(X \mid \bar{\lambda}) \geq p(X \mid \lambda)$

Ovo je iterativni proces u kom se novi model tekuće iteracije smatra početnim modelom naredne iteracije i kompletan proces se ponavlja dok se ne dostigne određen prag konvergencije. U sistemima za prepoznavanje govornika koji su bazirani na GMM modelima, javlja se model sveta (eng. world model), koji se još naziva i univerzalni pozadinski model (eng. Universal Background Model) ili UBM. Ovaj model se podvrgava treningu koristeći prethodno opisani EM algoritam. UBM model predstavlja raspodelu vektora karakteristika koja je nezavisna od govornika. Pri upisivanju novog govornika u sistem, parametri pozadinskog modela se prilagođavaju raspodeli karakteristika novog govornika. Adaptirani model se zatim koristi kao model tog govornika. Na ovaj način se model govornika ne pravi od nule, već se koriste prethodno poznate informacije.

Skladištenje podataka se odvija preko standardne baze podataka preko SQL upita. Predlaže se interektivni sistemski pristup bazi podataka u kojoj se govor prepoznaju pomoću programa za prepoznavanja glasa, a prepoznat tekst se pretvara u tekst, a zatim se pretvara u standardan SQL upit.



Slika 31 Dijagram skladištenja podataka

Prvi modul (automatsko prepoznavanje glasa) uzima glas korisnika i unosi odgovarajući tekst kao izlaz po uzorku na Youa (You, Lu, & Shixing, 2009). To je kontinualni govorni signal koji je nezavisan od govornika. Četiri programa se koriste za implementaciju ovog modula. Glas korisnika se uzima kao ulaz snimljen pomoću mikrofona i čuva se kao .wav datoteka. Ova datoteka se uzima kao ulaz za obradu podataka. Konfiguracioni fajl sadrži potrebne detalje za obradu wav datoteke, brzina uzorkovanja (Gada, Rao, & Samudravijaya, 2013) za obradu govornog signala koji se pohranjuje u sam fajl.Svakako se prvo izbriše buka objašnjena u prethodnim poglavljima.

Sledeći modul je modul koji konvertuje ulazni uzorak u tekst. Prvo se napiše xml konfiguracija za kreiranu bazu podataka. Odvajaju se reči u rečenicama, pregledaju se veznici u rečenicama npr. i, ili, za (and, or, for...). Samim tim se kreiraju atributi tabela.

U poslednjem modulu se razmatraju sql upiti koji su dobijeni preko ulaznog govornog uzorka, te se tim putem pronalaze potrebni podaci iz baze podataka.

Trenutno u ovoj bazi postoji 540 fajlova po jednoj osobi, a pošto je snimljeno ukupno 30 osoba, dalje implicira da je u bazi ukupno 16200 fajlova (bez razdvajanja reči iz rečenica).

6.7. FORMIRANJE EKSPERIMENTALNOG LABORATORIJSKOG SISTEMA SA VIŠE INSTANCI SISTEMA ZA PREPOZNAVANJE GOVORNIKA

Cilj prepoznavanja govornika je jednostavniji ako se ispitanici razlikuju (Milošević, 2020). Na osnovu karateristika govornika, najpogodniji scenario bi bio između veće grupe ispitanika jer postoje velike varijacije između njih, dok su kod jednog govornika male. Podela koju su izvršili Hansen i Hasan (Hansen & Hasan, 2015) se odnosi na to da uzroci varijabilnosti glasa jednog istog govornika mogu dolaziti od govornika lično, razgovora ili tehnologije koja obraduje signal govora (Zheng & Li, 2017).

Tokom procesa snimanja potrebno je izdvojiti karakteristike analizom spektra signala, koje poseduju fonetsku informaciju koja je u vezi sa sadržajem govora, kao i informaciju o identitetu govornika. U ovom radu su obrađeni MFCC koeficijenti, visina tona, formanti, kao i energija zvučnog signala.

Fast Fourier transformacija pomaže da se dobije cepstrum[21] signala. Korišćenjem Melove skale (skala prilagođena ljudskoj u kombinaciji sa kepstralnom analizom dobiva se Melov frekvencijski kepstrum. Koeficijenti koji sacinjavaju MFC nazivaju se MFCC koeficijenti.

6.8. OBUKA INSTANCI SISTEMA ZA PREPOZNAVANJE GOVORNIKA

Na početku samog projekta, moraju da se unesu neki glasovni zapisi, te je snimljeno 30 osoba između 18 i 35 godina, koje su izgovarali reči po tekstualnim uputstvima na srpskom i engleskom jeziku. Izjave su se snimale u kratkim (do 1 sekunde), srednjim (od 1 do 2 sekunde) i dugim zapisima (preko 2 sekunde). Pomoću tih reči/rečenica, model se obučavao. Svaki govorni signal treba da bude obeležen (snimljena kao labela) na način koji nas asocira na tekst koji je sadržan u njemu.

[21] "cepstrum" pretvara signal u frekventni domen kroz Furijeovu transformaciju, uzima logaritam, i zatim primenjuje drugu Furijeovu tranformaciju. Ovo naglašava harmonijsku strukturu originalnog spektra.

6.9. REALIZACIJA SCENARIJA MOGUĆIH VARIJACIJA U INTERAKCIJI U CILJU ISPITIVANJA EKSPLOATACIONIH PERFORMANSI SISTEMA ZA PREPOZNAVANJE GOVORNIKA

U ovom radu, dorađena su četiri programa otvorenog koda za prepoznavanje govornika. U narednom tekstu biće šire prikazana izrada scenarija mogućih varijacija u interakciji sistema za prepoznavanja govornika iz poglavlja 5.3.

```
def prepoznaj_govornika(govornik, mikrofon):
    """Transkripcija govora snimljenog sa mikrofona`.

    "uspešno": a boolean indicating whether or not the API request was
               successful
    "greška":   `None` if no error occured, otherwise a string containing
               an error message if the API could not be reached or
               speech was unrecognizable
    "transkripcija": `None` if speech could not be transcribed,
               otherwise a string containing the transcribed text
    """
    # proveri govornika i tip mikrofona
    if not isinstance(recognizer, sr.Recognizer):
        raise TypeError("`govornik` must be `govornik` instance")

    if not isinstance(microphone, sr.Microphone):
        raise TypeError("`mikrofon` must be `mikrofon` instance")

    # prilagodi osetljivost govornika na ambijentalnu buku i snimite zvuk
    # sa mikrofona
    with mikrofon as source:
        recognizer.adjust_for_ambient_noise(source) # #  analiziraj izvor zvuka u trajanju od 1 sekunde
        audio = recognizer.listen(source)

    # odgovor
    response = {
        "uspešno": True,
        "greška": None,
        "transkripcija": None
    }

    # pokušaj prepoznati govor na snimku
    # ako postoji izuzetak RequestError or UnknownValueError,
    #   uskladi odgovor
    try:
        response["transkripcija"] = recognizer.recognize(audio)
    except sr.RequestError:
        # API ne reaguje
        response["uspešno"] = False
        response["greška"] = "API ne reaguje"
    except sr.UnknownValueError:
        # speech was unintelligible
        response["greška"] = "Nije moguće prepoznati govornika"

    return response
```

Slika 32 Prepoznavanje govornika (programski kod)

Za sva tri scenarija, kao i za ostale koji nisu navedeni, procedura je ista. Prepoznavanje govornika se vrši kroz bazu, a takođe se vrši i prepoznavanje uređaja putem kojeg se unosi glas (slika 32).

Takođe, prepoznavanje jezika se vrši na sledeći način (slika 33).

```
import speech_recognition as sr

r = sr.Recognizer()

with sr.AudioFile('path/to/audiofile.wav') as source:
    audio = r.record(source)

r.recognize(audio, language='srp-<SRP')
```

Slika 33 Prepoznavanje jezika govornika

6.10. PROGRAMIRANJE IZVRŠNIH KONFIGURACIJA NA OSNOVU SCENARIJA

Termin otvoreni kod označava programsko rešenje čiji je prvobitni raspoloživ za slobodno korišćenje ili modifikaciju svim programerima kako god oni to smatrali ispravno (Radmanić, 2019). Za razliku od vlasničkog softvera, softver otvorenog koda je računarski softver koji je razvijen kao javna, otvorena saradnja i dostupan javnosti. Uopštena procedura za sve programe otvorenog koda koji se koriste u ovom radu je:

```
var mojRezultat = prepoznavanjeGovornika.rezultat;

prepoznavanjeGovornika.rezultat = false;

...

var gramatika = 'C:\Users\Korisnik\Desktop\Doktorat\Recenice.txt;'

var prepoznavanje = new prepoznavanjeGovornika();

var listaPrepoznavanjaGovornika = new listaGramatika();

listaPrepoznavanjaGovornika.addFromString(gramatika, 1);
```

```
prepoznavanje.gramatika = listaPrepoznavanjaGovornika;

//prepoznavanje.continuous = false;

prepoznavanje.lang = 'en-US';

prepoznavanje.interimResults = false;

prepoznavanje.maxAlternatives = 1;
```

Ovakva procedura nam pokazuje jasne ciljeve programa otvorenog koda. Prvo se kreira skladište uzoraka (reči/rečenica) pri unosu putem različitih mikrofona (telefon, integrisani laptop mikrofon, eksterni mikrofon). Uprkos napretku koji je postignut u automatskom učenju gramatike, u ova četiri programa je ubačena ručno ili delimično ručno napravljena gramatika za prihvatljivu preciznost.

- SPEAR

SPEAR je komplet alata za prepoznavanje govornika zasnovan na Bobu, dizajniran za pokretanje eksperimenata verifikacije/prepoznavanja govornika (Curatelli, Mayora-Ibarra, & Carotenuto, 1999). SPEAR je dizajniran tako da bi trebalo da bude lako izvoditi eksperimente kombinovanjem narednih alata:

- Baze podataka za prepoznavanje govornika i njihovi protokoli,
- Detekcija glasovne aktivnosti,
- Svojstva ekstrakcije
- Alati za prepoznavanje/verifikaciju.

Alati u SPEAR-u su dizajnirani u sedam faza:

1) Otkrivanje glasovne aktivnosti
2) Ekstrakcija karakteristika
3) UBM Obuka i projekcija (izračunavanje dovoljne statistike)
4) Podprostorna obuka i projekcija (za ISV, JFA i I-Vektorsko modeliranje)
5) Uslovljavanje i kompenzacija (za I-Vektorsko modeliranje)
6) Upis
7) Normalizacija rezultata

Otkrivanje glasovne aktivnosti ima za cilj da filtrira deo koji nije govor.

Što se tiče ekstrakcije podataka, samo ime mu kaže da služi tome da izdvoji karakteristike. Tu dolaze u obzir:

- LFCC/MFCC ekstrakcija karakteristika,
- Ekstrakcija spektrograma,
- Normalizacija karakteristika,
- HTK čitač funkcija i
- SPro Feature reader.

Treći korak ima za cilj izračunavanje univerzalnog pozadinskog modela (UBM, eng. Universal Background Model) koji se naziva projektor. Zatim, izračunavanje dovoljne statistike u SPEAR-u je korak projekcije-ubm. Ima za cilj da projektuje kepstralne karakteristike koristeći prethodno obučeni projektor.

Podprostorna obuka i projekcija služe za procenu podprostora potrebnih za ISV, JFA i I-Vector. Na projekciju ovde upućuje ili projection-isv, projection-jfa ili projection-ivector. Ovde vidimo da je proces projekcije I-vektora ekstrakcija i-vektora.

Uslovljavanje i kompenzacija se odnosi na korišćenje alata I-Vector. Uključuje izbeljivanje, normalizaciju dužine, LDA i VCCN projekciju. Izbeljivanje se odnosi na metodu poravnanja spektra koja se neretko koristi kod digitalne obrade slike i zvuka (Vaseghi, 2008) (Grozdić, 2017).

Upis definiše fazu u kojoj se nekoliko (projektovanih ili kompenzovanih) karakteristika jednog identiteta koristi za upis modela za taj identitet. U najlakšem slučaju, karakteristike se jednostavno usredsređuju, a prosečna karakteristika se koristi kao model.

U završnoj fazi normalizacije rezultata, modeli se upoređuju sa karakteristikama i izračunava se rezultat sličnosti za svaki par modela. Neki od modela (tzv. T-Norm-Model) i neke karakteristike sonde (tzv. Z-Norm-sonde-karakteristike) su podeljeni, tako da se mogu koristiti za normalizaciju rezultata kasnije.

U svakom slučaju, rezultati ovih eksperimenata se direktno upoređuju kada se koristi isti skup podataka (slika 25).

```
#!/usr/bin/env python

import bob.db.multipie
import voicereclib

multipie_voice_directory = "[YOUR_MULTI-PIE_VOICE_DIRECTORY]"
multipie_annotation_directory = "[YOUR_MULTI-PIE_ANNOTATION_DIRECTORY]"


mobilemic = ('02_0', '05_0', '16_0', '01_0', '04_1', '05_0', '05_1', '14_0', '13_0', '08_0', '09_0', '12_
0', '11_0')

database = voicereclib.databases.DatabaseSpearBobOlja(


    database = bob.db.multipie.Database(

        original_directory = multipie_image_directory,

        annotation_directory = multipie_annotation_directory

    ),

    name = "multipie",


    all_files_options = {'mobilemic' : mobilemic},

    extractor_training_options = {'mobilemic' : mobilemic},

    projector_training_options = {'mobilemic' : mobilemic, 'upis': 3, 'test': True},

    enroller_training_options = {'mobilemic' : mobilemic}
```

Slika 34 Spear konfiguracija za prepoznavanje govornika putem mobilnog telefona

- MARF

Centar MARF-a predstavlja klasa MARF. Ona je glavni server i mesto na kom su smeštena sva podešavanja. Takođe, ova klasa sadrži glavne metode, koje su sastavni deo tipičnog procesa prepoznavanja glasa. To su metode soPreprocessing, soFeatureExtraction i soClassification.

Kod MARF-a, izabrana je konfiguracija za korake preprocesiranja, izdvajanja karakteristika i klasifikacije. Kao što je prethodno navedeno, svaki sistem za prepoznavanje glasa se sastoji iz faze

upisa (treninga) i faze testiranja (prepoznavanja). Tako je i u kreiranju aplikacije na MARF platformi. Trening se izvršava da bismo dobili modele osoba koje će se nalaziti u sistemu, dok je prepoznavanje konkretan proces identifikacije upoređivanjem uzorka sa prethodno sačuvanim modelima tokom faze treninga.

Za učitavanje uzoraka koriste se metode glavne klase MARF, *setSampleFormat(), setSamplesDir()* i *setSampleFile()*. Koristeći *setSampleFormat()* podešavamo format ulazne datoteke tj. uzorka. Pozivanjem metode *setSamplesDir()* postavlja se direktorijum iz koga će se učitati uzorci, dok metodom *setSampleFile()* učitavamo samo jedan uzorak. Ove metode koriste klasu *SampleLoader* i njoj pripadajuće metode. Moguće je definisati neku drugu klasu za učitavanje uzoraka, što se čini metodom *setSampleLoaderPluginClass()*.

U MARF-u preprocesiranje je predstavljeno klasom *Preprocessing* i identičnim interfejsom *IPreprocessing*. Ova klasa ima polje oSample koje predstavlja uzorak nad kojim se vrši preprocesiranje. Pozivanjem metode *removeNoise()* izvršava se uklanjanje šuma sa oSample objekta korišćenjem low-pass FFT filtera. Metoda *removeSilence()* analizira uzorak i određuje u kojim trenucima je amplituda ispod određenog nivoa, a zatim te delove izbacuje iz objekta oSample. Nakon što smo izvršili ove dve transformacije, pozivanjem metode *normalize()* vršimo normalizaciju objekta oSample, nakon čega je ovaj objekat spreman za korak izdvajanja karakteristika.

U MARF-u je implementirano dve metode za izdvajanje karakteristika LPC i FFT. Metoda izvlačenja cepstralnih koeficijenata je predviđena, međutim i dalje nije implementirana. Pri programiranju naše aplikacije, dostupna nam je metoda *setFeatureExtractionMethod()* glavne klase MARF, čijim korišćenjem vrlo jednostavno možemo podesiti neku od implementiranih metoda.

Za fazu klasifikacije omogućeno je nekoliko metoda, a između ostalih različite vrste odstupanja kao što su ranije opisane Chebyshev i Euklidsko. Pored odstupanja postoji implementirana je i metoda neuronskih mreža, dok za neke stohastičke metode i HMM modele postoji podrška, ali bez implementacije. Metoda glavne klase MARF *setClassificationMethod()* je dostupna za korišćenje prilikom programiranja aplikacije.

Da bismo različite trening uzorke povezali sa odgovarajućim osobama, koristimo jedan tekstualni fajl u CSV formatu, u kojem u svakom redu definišemo šifru jedne osobe, njeno ime i prezime i spisak trening uzoraka koji pripadaju toj osobi. Jedna linija u ovom fajlu izgleda ovako:

ID osobe, Ime osobe, trening-uzorak1.wav|trening-uzorak2.wav

Jedna osoba je predstavljena jednom linijom ovog tekstualnog fajla. Fajl korišćen u primeru je nazvan *govornici.txt* i jedan njegov deo izgleda ovako:

```
0,Nepoznata Osoba

1,Osoba 1,Osoba1-1.wav|Osoba1-2.wav|Osoba1-3.wav|Osoba1-4.wav| Osoba1-
5.wav|Osoba1-6.wav|Osoba1-7.wav|Osoba1-8.wav|Osoba1-9.wav|Osoba1-
10.wav|Osoba1-11.wav| Osoba1-12.wav|Osoba1-13.wav|Osoba1-15.wav|Osoba1-
16.wav|Osoba1-17.wav|Osoba1-18.wav|Osoba1-19.wav|Osoba1-20.wav|Osoba1-
21.wav|Osoba1-22.wav|Osoba1-23.wav|Osoba1-24.wav|Osoba1-25.wav|Osoba1-
26.wav|Osoba1-27.wav|Osoba1-28.wav|Osoba1-29.wav|Osoba1-30.wav

...
```

SpeakerIdentDb klasa predstavlja pomoćnu klasu u našoj aplikaciji. I pored toga, njena uloga je veoma bitna jer sadrži metode koje služe za rad sa govornici.txt fajlom. Uz pomoć metoda ove klase možemo da izvučemo podatke iz samog fajla i popunimo interne strukture podataka na nivou same aplikacije. Takođe, ova klasa služi za beleženje statistike prepoznavanja, kao i njeno resetovanje. Bitna polja ove klase su *oDB* i *oConnection*. ***oDB*** predstavlja hash tabelu govornika iz baze tj. tekstualne datoteke, a **oConnection** je konekcija ka bazi tj. otvorena tekstualna datoteka. Metode ove klase su ***connect(), query(), close(), getIDByFilename(), getName(), addStats(), resetStats()*** *i* ***printStats().***

Jedna od bitnijih metoda (slika 26) je *ident()* koja služi za identifikaciju govornika koristeći MARF konfiguraciju, audio datoteku wav formata i opciono šifru očekivanog govornika. Kao ulazne argumente ova metoda prima String koji predstavlja konfiguraciju koja se koristi, String koji predstavlja naziv audio datoteke u wav formatu i ceo broj koji predstavlja šifru očekivanog govornika. Ova metoda izgleda ovako:

```
public static final void ident(String pstrConfig, String pstrFilename, int piExpectedID)
    throws MARFException {
        /*
         * Ukoliko nije definisana ocekivana sifra osobe
         * pokusaj da pronadjes na osnovu imena datoteke
         */
        if(piExpectedID < 0) {
            piExpectedID = bazaConf.getIDByFilename(pstrFilename, false);
        }

        int iIdentifiedID = -1;
        // Drugi najbolji pogodak
        int iSecondClosestID = -1;

        long lTimeStart = System.currentTimeMillis();
        {
            MARF.setSampleFile(pstrFilename);
            MARF.recognize();

            // Prvi pogodak
            iIdentifiedID = MARF.queryResultID();

            // Drugi najbolji pogodak
            iSecondClosestID = MARF.getResultSet().getSecondClosestID();
        }
        long lTimeFinish = System.currentTimeMillis();

        int iOneSecInMs  = 1000;
        int iOneMinInMs  = 60 * iOneSecInMs;
        int iOneHourInMs = 60 * iOneMinInMs;
        int iOneDayInMs  = 24 * iOneHourInMs;
        long lDuration = lTimeFinish - lTimeStart;
        long lDays = lDuration / iOneDayInMs;
        long lModTimeLeft = lDuration % iOneDayInMs;
        long lHours = lModTimeLeft / iOneHourInMs;
        lModTimeLeft %= iOneHourInMs;

        long lMinutes = lModTimeLeft / iOneMinInMs;
        lModTimeLeft %= iOneMinInMs;
        long lSeconds = lModTimeLeft / iOneSecInMs;
        lModTimeLeft %= iOneSecInMs;
        long lMilliSeconds = lModTimeLeft;

        StringBuffer oDuration = new StringBuffer()
            .append(lDays).append("d:")
            .append(lHours).append("h:")
            .append(lMinutes).append("m:")
            .append(lSeconds).append("s:")
            .append(lMilliSeconds).append("ms:")
            .append(lDuration).append("ms");

        System.out.println( "               File: " + pstrFilename);
        System.out.println( "             Config: " + pstrConfig);
        System.out.println( "      Processing time: " + oDuration);
        System.out.println( "         Speaker's ID: " + iIdentifiedID);
        System.out.println( "   Speaker identified: " +       bazaConf.getName(iIdentifiedID));
```

Slika 35 MARF - metoda ident()

Svakako najbitnija je *main()* metoda koja je prikazana na slikama 27-29. U okviru ove metode vrši se učitavanje podataka iz tekstualne datoteke. Kada je oformljena baza osoba u radnoj memoriji, potrebno je podesiti željene algoritme za sve faze u procesu prepoznavanja glasa. Za svaku od faza postoji posebna metoda:

1. Preprocesiranje - **MARF.setPreprocessingMethod(**MARF.DUMMY**)**
2. Izdvajanje karakteristika - **MARF.setFeatureExtractionMethod(**MARF.LPC**)**
3. Klasifikacija - **MARF.setClassificationMethod(**MARF.EUCLIDEAN_DISTANCE**)**

```
public static void main(String[] args) {
        try {
            try {

                // Ucitavanje podataka iz tekstualne datoteke i popunjavanje internih
                // struktura podataka
                bazaConf.connect();
                bazaConf.query();

                bazaSpeak.connect();
                bazaSpeak.query();

            } catch (StorageException ex) {
                System.out.println("Neuspesno ucitavanje podataka!");
            }


            // Podesavanje algoritama za proces prepoznavanja glasa
            // - koristimo WAV format uzoraka
            // - U preprocesiranju samo normalizaciju
            // - Za izdvajanje karakteristika LPC algoritam
            // - Za klasifikaciju Euklidsko odstojanje
            MARF.setSampleFormat(MARF.WAV);
            MARF.setPreprocessingMethod(MARF.DUMMY);
            MARF.setFeatureExtractionMethod(MARF.LPC);
            MARF.setClassificationMethod(MARF.EUCLIDEAN_DISTANCE);
```

Slika 36 MARF - main() metoda_1.deo

```
// Omogucavanje debugging moda
        Debug.sbDebugOn = true;

        // Ucitavanje fajlova u sistem
        File[] aoSampleFiles = new File("olja-train").listFiles();
        Debug.debug("Files array: " + aoSampleFiles);
        if (Debug.isDebugOn()) {
            System.getProperties().list(System.err);
        }
        String strFileName = "";

        // Vrsimo treniranje na osnovu svih trening uzoraka
        for (int i = 0; i < aoSampleFiles.length; i++) {
            strFileName = aoSampleFiles[i].getPath();
            if (aoSampleFiles[i].isFile() && strFileName.toLowerCase().endsWith(".wav")) {
                try {
                    train(strFileName);
                } catch (MARFException ex) {
                    Logger.getLogger(MojSpeakerIdentApp.class.getName()).log(Level.SEVERE, null, ex);
                }
            }
        }

        System.out.println("Done training on folder \" olja-train \".");
```

Slika 37 MARF - main() metoda_2.deo

```
System.out.println("Starting test on folder \" olja-test \".");

        String strConfig = MARF.getConfig();

        // Izlistavanje svih datoteka u test direktorijumu
        aoSampleFiles = new File("olja-test").listFiles();

        // Za svaki uzorak iz test direktorijuma vrsi se identifikacija
        for(int i = 0; i < aoSampleFiles.length; i++){

            strFileName = aoSampleFiles[i].getPath();

            if(aoSampleFiles[i].isFile() && strFileName.toLowerCase().endsWith(".wav")){
        ident(strConfig, strFileName, 3);
    }
        }
```

Slika 38 MARF - main() metoda_3.deo

- ALIZE

ALIZE Core biblioteka čini osnovu ALIZE platforme, platforme otvorenog koda za prepoznavanje govornika. Ovde je fokus na funkcijama niskog nivoa, uglavnom na svim funkcijama koje su potrebne za korišćenje GMM, kao i na I/O funkcije za različite formate datoteka. To je osnova na kojoj se mogu implementirati različiti algoritmi za prepoznavanje govornika.

Alati koji su primenjeni u istraživanju su:

- EnergyDetector – Detektor govorne aktivnosti
- NormFeat – Normalizacija (različiti režimi)
- TrainWorld – procena UBM korišćenjem
- TrainTarget – Train modeli

Opšta arhitektura kompleta alata koji se nalaze u Alize platformi je zasnovana na podeli funkcionalnosti između nekoliko softverskih servera. Glavni serveri su server funkcija, koji upravlja akustičnim podacima, server mešavine koji se bavi modelima (čuvanje, modifikacija, vezivanje komponenti, čuvanje/čitanje...), i server statistike koji implementira sva statistička izračunavanja (Statističke procene zasnovane na EM, izračunavanje verovatnoće, itd.)

Fragment programskog koda za prepoznavanje govornika u Alize tehnologiji otvorenog koda je prikazan na slici 30. Korisnički kod predstavlja istu strukturu, organizovanu između glavnih

servera, pomažući razvoj i razumevanje izvornog koda. Što se tiče arhitekture softvera ona omogućava laku distribuciju servera na različite niti ili računare.

```
import AlizeSpkRec.*;
import AlizeSpkRec.SimpleSpkDetSystem.SpkRecResult;

public class PrepoznavanjeGovornikaAlize extends PrepoznavanjeGovornika {

    private void displayStatus(SimpleSpkDetSystem spkDetSystem) throws Exception {
        Log.i("ALIZE","*************************************************");
        Log.i("ALIZE","Status:");
        Log.i("ALIZE","  # karakteristike: " + spkDetSystem.featureCount());
        Log.i("ALIZE","  # modeli: " + spkDetSystem.speakerCount());
        Log.i("ALIZE","  UBM je učitan: " + spkDetSystem.isUBMLoaded());
        Log.i("ALIZE","*************************************************");
```

Slika 39 Prikaz fragmenta programskog koda za prepoznavanje govornika u Alize tehnologiji otvorenog koda

- HTK

Glavni koraci u alatu HTK su sledeći:

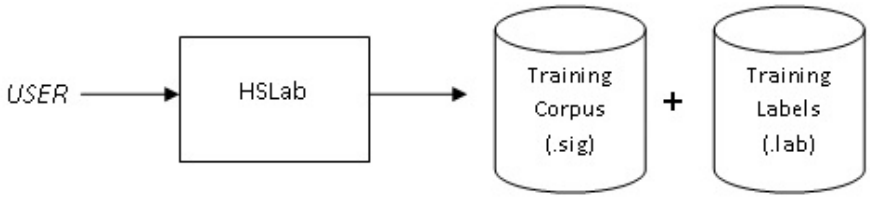
- Stvaranje baze podataka treninga: svaki element reči je zabeležen nekoliko puta i snimljen je kao labela pod određenim nazivom;
- Akustična analiza: obuka talasa se konvertuje u nekoliko serija vektora koeficijenata;
- Definisanje modela: prototip skrivenog Markovljevog modela je definisan za svaki element rečnika;
- Obuka modela: svaki skriveni Markovljev model je inicijalizovan i obučavan sa podacima za obuku;
- Gramatika prepoznavanja (šta će prepoznati) je definisana;
- Prepoznavanje nepoznatog ulaznog zapisa;

Strukture direktorijuma su organizovane na sledeći način:

- ***data/*** : služi za skladištenje obuka (treninga) i test podataka (glasovni zapis, labele, itd.) sa po 2 poddirektorijuma ***data/train/*** i ***data/test/*** da bi se razdvojili podaci koji su korišćeni za obuku od onih koji se koriste za procenu performansi;
- ***analysis/*** : služi za skladištenje datoteka koje se tiču akustičke analize;
- ***training/*** : služi da se skladište datoteke koje se tiču inicijalizacije i obuke;
- ***model/*** : za skladištenje modela za prepoznavanje (HMM);
- ***def/*** : za skladištenje datoteka koje se odnose na definisanje zadatka;
- ***test/*** : za skladištenje datoteka koje se tiču ispitivanja, odnosno testiranja.

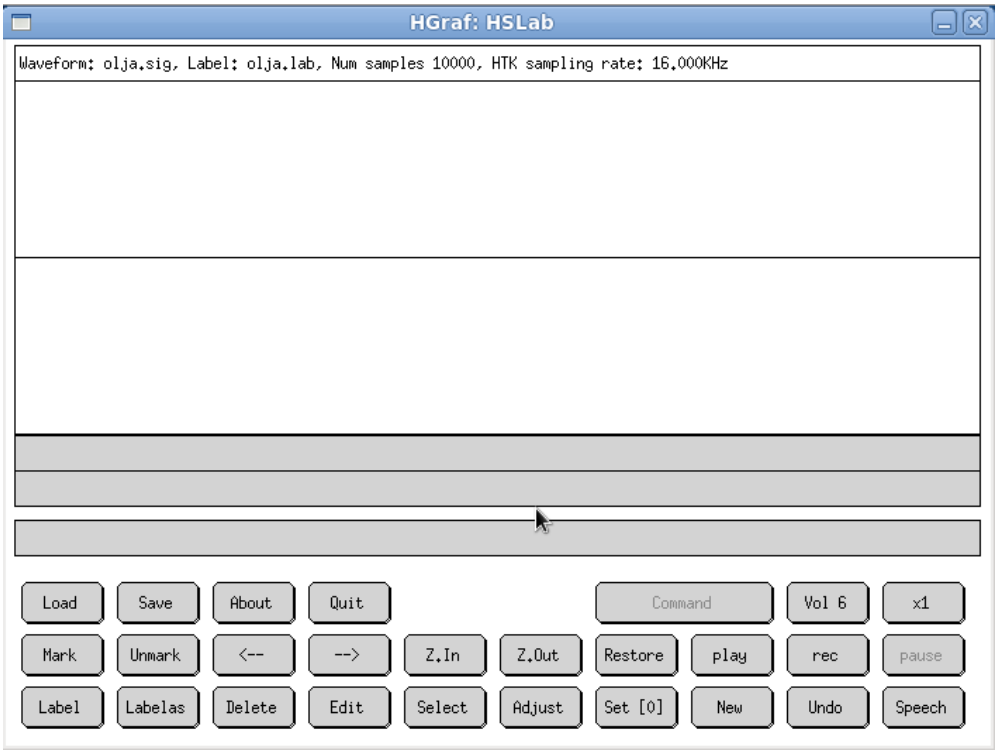
Neke od standardnih opcija koje sadrži svaki HTK alat su:

- -A : prikazuje komandnu liniju argumenata;
- -D : prikazuje podešavanja konfiguracije;
- -T 1 : prikazuje neke informacije o algoritmu.



Slika 40 Snimanje i obeležavanje glasovnog zapisa u HTK alatu

Ako se glasovni uzorak unosi direktno preko HTK alata, onda koristimo funkciju HSLab, odnosno prvo pišemo putanju u command prompt-u do HTK alata (*htkInstalacija/bin/HSLab olja.sig*), te nakon toga pritisnemo dugme „Rec“ koje služi za snimanje i „Stop“ kada želimo da zaustavimo snimanje.



Slika 41 Funkcija za snimanje u HTK alatu – HSLab

Prepoznavanje određenu osobe na osnovu glasovnog zapisa u ovom alatu se vrši kroz sledeće procedure:

- Ulazni govorni signal *input.sig* se prvo pretvara u niz akustičkih vektora (u ovom slučaju MFCC) sa komandom *HCopy*, na isti način kao što je bilo i sa podacima za obuku (trening). Rezultat je *input.mfcc* fajl, koji se najčešće naziva akustičko posmatranje.

- Zatim, imamo ulazno posmatranje sa određenim algoritmom, gdje je rezultat ponovno prepoznavnje Markovljevih modela. Ovako bi to izgledalo sa komandom HVite:

**HVite -A -D -T 1 -H hmmsdef.mmf -i reco.mlf -w net.slf **
dict.txt hmmlist.txt input.mfcc

-A prikazuje elemente komandne linije.

-D prikazuje konfiguraciona podešavanja.

-T prikazuje informacije o akcijama algoritma.

input.mfcc je ulazni podatak za prepoznavanje.

hmmlist.txt je lista imena modela koji se koriste (olja, boris, igor, jovana, bojan, marko... i sil).

Sil predstavlja tišinu (objašnjena u daljem tekstu).

dict.txt predstavlja rečnik.

net.slf se odnosi na mrežu.

reco.mlf je izlazno prepoznavanje transkripcionog fajla.

Hmmsdef.mmf sadržava defniciju HMM-a. Moguće je ponoviti *–H* opciju, koja bi izgledala u ovom slučaju:

-H model/hmm1/hmm_olja

-H model/hmm1/hmm_igor

-H model/hmm1/hmm_boris

-H model/hmm1/hmm_jovana

-H model/hmm1/hmm_bojan

-H model/hmm1/hmm_marko

...

-H model/hmm1/hmm_sil

Jedan od interaktivnijih načina testiranja prepoznavanja glasa je to da se koristi „direktni ulaz“, kao opcija HVite, koja izgleda ovako:

**HVite -A -D -T 1 -C directin.conf -g H hmmsdef.mmf **
-w net.slf dict.txt hmmlist.txt

Ovako bi izgledao rezultat, nakon što izvršimo ovu komandu:

READY[1]>

Press return to start sampling

pa trebamo reći reči ili rečenice koje se nalaze u trening fazi ili samo da ćutimo što bi prepoznao kao „Sil“, odnosno tišinu, te bi ispisao ovakve nekakve rezultate:

START_SIL START_SIL START_SIL OLJA END_SIL END_SIL END_SIL = = [305] frames -54.4413 [Ac=-1660 4.6 LM=0.0] (Act=12.0)

Ili

START_SIL START_SIL START_SIL IVAN END_SIL END_SIL END_SIL = = [229] frames -55.7739 [Ac=-12764 4.6 LM=0.0] (Act=11.9)

6.11. EKSPERIMENTALNO ISPITIVANJE UTICAJA VARIJACIJA NA PERFORMANSE SISTEMA PREMA POSTAVLJENOM SCENARIJU

Cilj ovog rada je da ispita uticaj varijacija u interakciji korisnika na performanse sistema za prepoznavanje govornika. Efikasnost sistema se ogleda u tačnosti i brzini biometrijskih metoda. Tačnost se može kvantitavno prikazati merama FAR, FRR i EER. Veliku pažnju treba posvetiti i brzini kojom se vrši poređenje. Zbog velikog broja izvršenih poređenja, sporije metode se ne mogu koristiti za identifikaciju. Tačnost i brzina mogu biti izražena kao posebna svojstva. FAR je verovatnoća da sistem prihvati uzorak koji se ne poklapa sa uzorom u bazi. FRR je verovatnoća da sistem ne prepozna uzorak koji se poklapa sa uzorom u bazi. EER je stopa po kojoj su i stopa prihvatanja i odbijanja grešaka iste. Što je ova stopa manja, to je sistem pouzdaniji.

6.12. REZULTATI STATISTIČKE OBRADE PRIKUPLJENIH PODATAKA

Pre nego što predstavimo rezultate istraživanja, prvo ćemo predstaviti kratak pregled mogućeg načina procene performansi biometrijskog sistema.

Efikasnost biometrijskog sistema ogleda se u tačnosti i brzini donošenja odluka. Izračunavanje ocene sličnosti između dva navedena skupa funkcija osnova je za odlučivanje da li oba skupa podataka pripadaju istom entitetu ili ne. Ako je izračunati rezultat iznad izabrane vrednosti, nazvan prag *(eng. threshold),* verujemo da oba skupa funkcija pripadaju istom entitetu (Malik, Girdhar, Dahiya, & Sainarayanan, 2014). Onda može da se kaže da su oba skupa obeležja varijacije unutar klase.

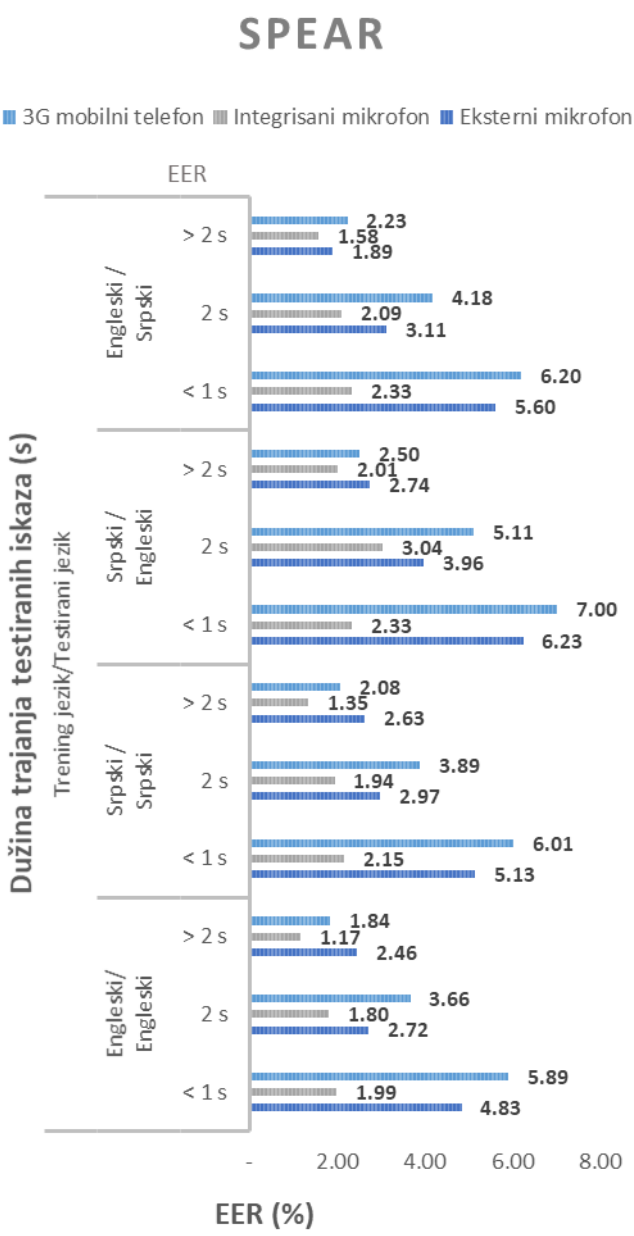
Naravno, veliku pažnju treba posvetiti i brzini postizanja rezultata. Zbog mnogih izvršenih poređenja, sporije metode ne bi trebalo koristiti u načinu identifikacije.

Prvo test istraživanje je jezička neusklađenost (srpski i engleski jezik) koje je rađeno u četiri tehnologije otvorenog koda za prepoznavanje govornika (SPEAR, MARF, ALIZE, HTK). Engleski i srpski su korišćeni kao različiti jezici u fazama upisa i prepoznavanja. Bilo je ukupno četiri scenarija koji su koristili korake za upis i prepoznavanje:

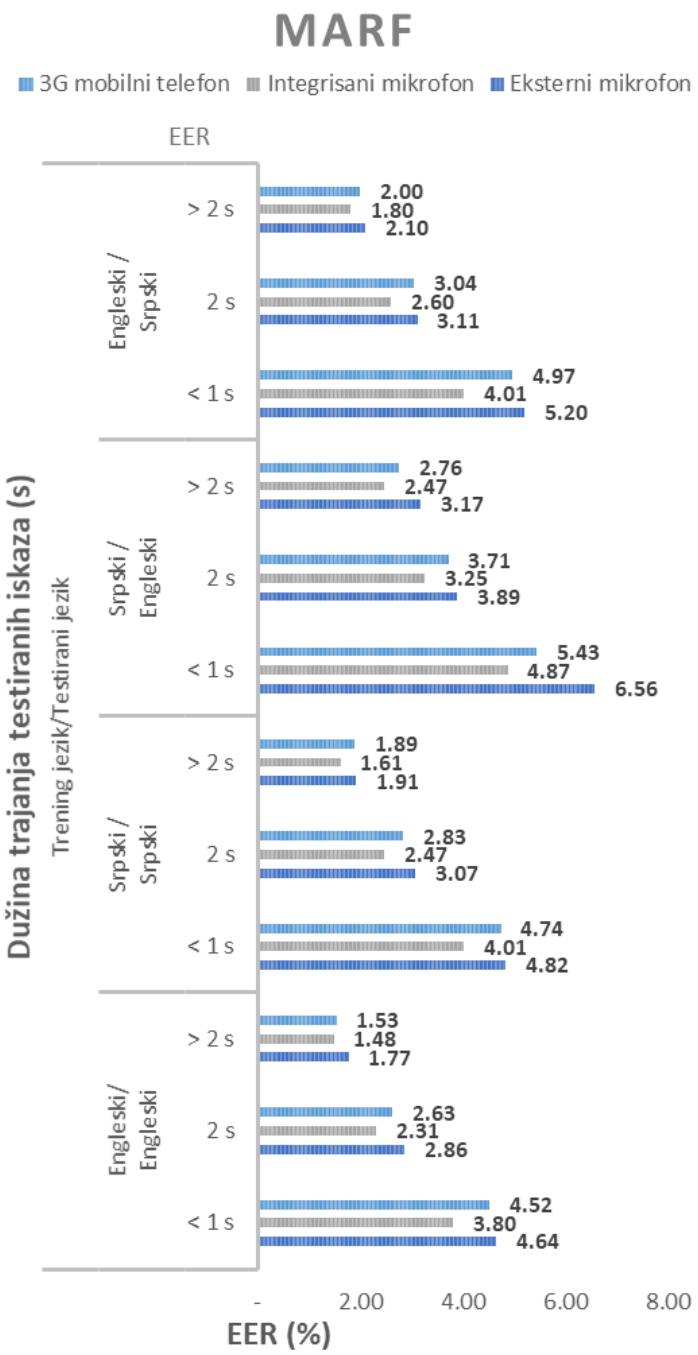
1. Engleski u obe faze;
2. Engleski pri upisu i srpski u fazi prepoznavanja;
3. Srpski pri upisu i engleski u fazi prepoznavanja;
4. Srpski u obe faze.

Za svaki scenario, merenja su vršena za devet različitih uslova rada, određena dužinom izgovorenog teksta (kratkim, srednjim i dugim) i korišćenim uređajem za snimanje (integrisani mikrofon, eksterni mikrofon i 3G mobilni telefon).

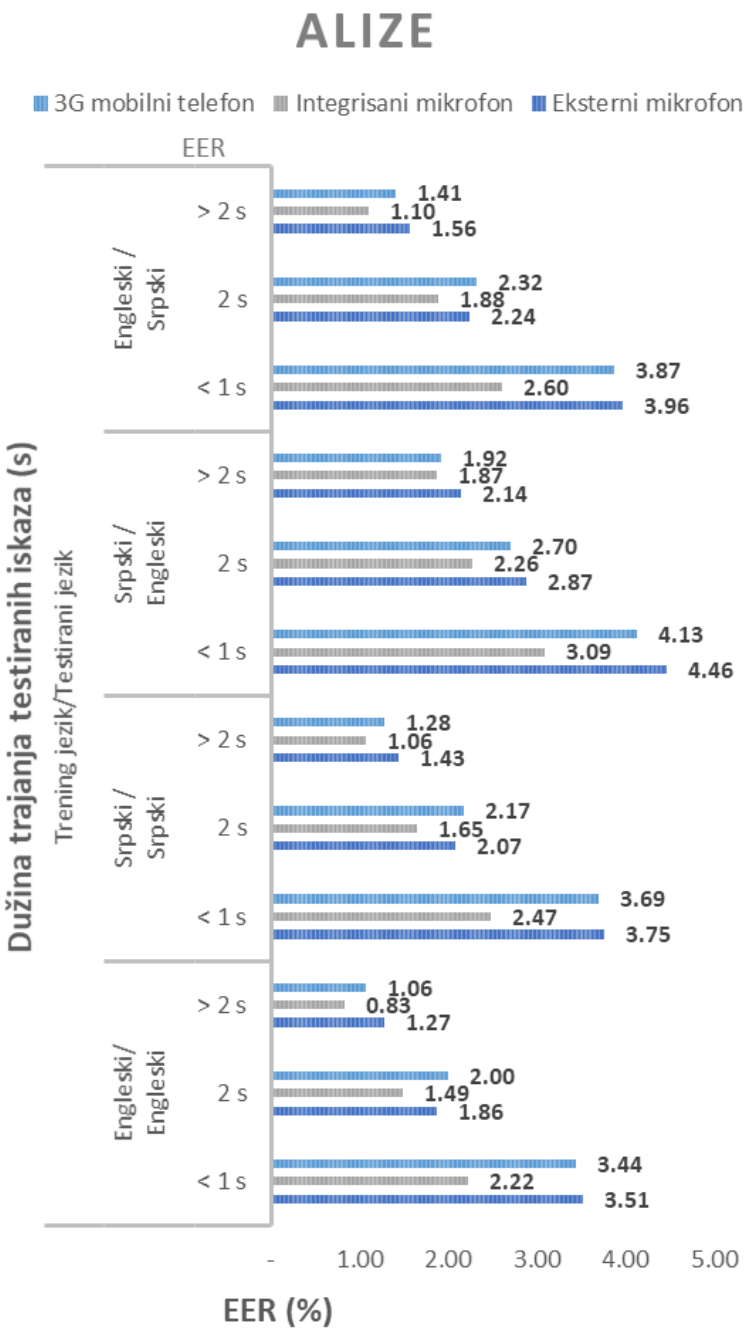
Performanse 144 testirane konfiguracije sa rešenjima otvorenog koda, merene jednakom stopom grešaka (EER), prikazane su na slikama 33-36.



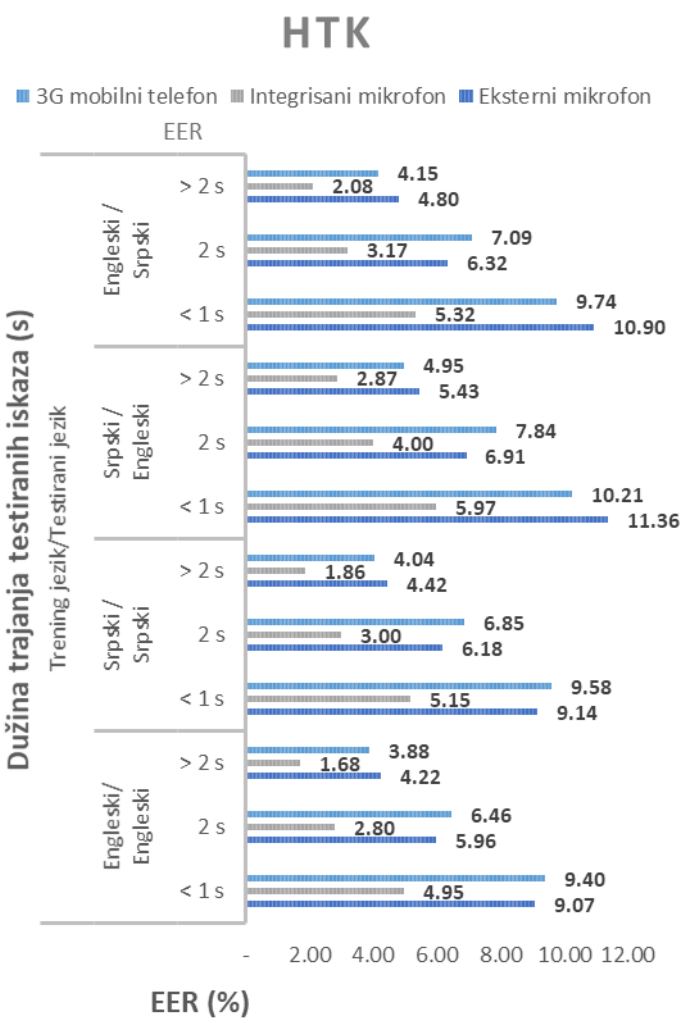
Slika 42 EER - Spear model prepoznavanja govornika



Slika 43 EER - MARF model prepoznavanja govornika



Slika 44 EER - ALIZE model prepoznavanja govornika



Slika 45 EER - HTK model prepoznavanja govornika

Najbolje performanse sistema za prepoznavanje govornika, grupisane prema rešenju otvorenog koda, prikazane su u tabelama 5–8. Svaka tabela prikazuje izmereni EER procenat za jedno rešenje i svaku jezičku kombinaciju.

Tabela 5 EER prikaz - SPEAR

SPEAR			
EER (%)		Test jezici	
		Srpski	*Engleski*
Trening jezici	*Srpski*	1.35	2.01
	Engleski	1.58	1.17

Tabela 6 EER prikaz - MARF

MARF			
EER (%)		**Test jezici**	
		Srpski	*Engleski*
Trening jezici	*Srpski*	1.61	2.47
	Engleski	1.80	1.48

Tabela 7 EER prikaz - ALIZE

ALIZE			
EER (%)		**Test jezici**	
		Srpski	*Engleski*
Trening jezici	*Srpski*	1.06	1.87
	Engleski	1.10	0.83

Tabela 8 EER prikaz - HTK

HTK			
EER (%)		**Test jezici**	
		Srpski	*Engleski*
Trening jezici	*Srpski*	1.86	2.87
	Engleski	2.08	1.68

7. EVALUACIJA DOBIJENIH REZULTATA

Evaluacija dobijenih rezultata istraživanja je završna faza realizacije istraživanja, a svrha joj je da zainteresovanoj grupi ljudi za biometriju glasa saopšti rezultate proučavanja sa dovoljno činjenica gde se može na kraju doneti kvalitetan zaključak.

Evaluacija i rangiranje dostupnih programskih rešenja se zasniva na merenju performansi sistema otvorenog koda u različitim uslovima interakcije korisnika sa sistemom. U ovoj fazi istraživanja prezentuje se ono što će ostati u okviru istraživačke dokumentacije.

U okviru predloženih metoda primenjene su nove linije kodova u četiri programa otvorenog koda u postupku akvizicije podataka u govoru. Za potrebe transformacije prikupljenih podataka govornika obrađeni su kroz tri modela uređaja, kroz dva jezika i kroz veličinu vremenskog uzorka za uticaj na krajnju preciznost u prepoznavanju.

Analiza rezultata pokazuje da su vrednosti parametra EER u prvom približavanju obrnuto proporcionalne dužini rečenica. Dodatna istraživanja trebala bi ukazati na optimalnu dužinu rečenice pri kojoj se pristupa minimalnom mogućem EER -u.

Testirana softverska rešenja su veoma osetljiva na karakteristike upisa putem mikrofona na bilo kojem uređaju, što u praksi može značajno uticati na performanse sistema za prepoznavanje ili verifikaciju govornika. Poželjno je koristiti rešenje koje u predprocesiranju minimizira uticaj varijacija u karakteristikama uređaja za snimanje govora. Osim toga, rezultati predstavljeni u ovom radu ukazuju na moguće probleme koji se mogu pojaviti pri lokalizaciji aplikacije za chatbot koja zahteva verifikaciju govornika.

Rezultati dobijeni sa ALIZE i SPEAR paketima su u skladu sa nalazima nedavnih istraživanja da i-vektori i rešenja zasnovana na PLDA i dalje predstavljaju dominantni pristupi prepoznavanju govornika za zadatke nezavisne od teksta (Alam, Bhattacharya, & Kenny, 2018).

7.1. INTERPRETACIJA REZULTATA DOBIJENIH U OKVIRU ISPITIVANIH PROGRAMSKIH SISTEMA ZA PREPOZNAVANJE GOVORNIKA SA OTVORENIM PROGRAMSKIM KODOM

Kada se koristi SPEAR, primetno je da korišćenje srpskog jezika pri upisu i engleskog u fazi testiranja rezultira nešto većim procentom EER -a. Slične konkluzije mogu se izvući za slučaj kada se koristi MARF.

Analizom svih dobijenih rezultata mogu se izvući sledeći zaključci u vezi sa odabranim rešenjima otvorenog koda:

- ALIZE rešenje otvorenog koda za prepoznavanje govornika pokazalo je najbolje performanse u svim jezičkim kombinacijama. Softverski alat za prepoznavanje govornika, SPEAR, došao je na drugo mesto, a MARF-ova platforma otvorenog koda na treće mesto. Najslabiji rezultati prepoznavanja govornika dobijeni su pomoću HMM Toolkit -a, HTK;
- Sva četiri testirana rešenja otvorenog koda dala su najbolje rezultate sa kombinacijom engleskog jezika kao jezika za obuku i za testiranje. Nešto slabiji rezultati dobijeni su u slučajevima kada se srpski koristio kao jezik za obuku i testiranje. U ovim slučajevima, EER se povećao sa 9% za MARF,a na 28% za ALIZE.

7.2. RANGIRANJE ISPITIVANIH PROGRAMSKIH SISTEMA ZA PREPOZNAVANJE GOVORNIKA SA OTVORENIM PROGRAMSKIM KODOM

Zanimljivo je napomenuti da su, u slučaju jezičke neusklađenosti za obuku i jezika za testiranje, izmereni bolji rezultati za kombinaciju engleskog kao jezika za obuku i srpskog jezika za testiranje. EER vrednosti za konfiguraciju srpskog jezika kao jezika za obuku i engleskog jezika za testiranje porasle su sa 27% za SPEAR, a na 70% za ALIZE, u poređenju sa kombinacijom engleskog kao jezika za obuku i srpskog kao testnog jezika.

U svim slučajevima, bez obzira na rešenje otvorenog koda koje se koristi (jezik i uređaj za snimanje glasa) vrednosti ERR se monotono smanjuju sa dužinom ispitne sekvence. Stoga se najbolje performanse sistema za prepoznavanje govornika postižu za najduži niz ispitivanja. Proširivanjem

rečenica iz kratkih zapisa (do 1 s) na duge zapise (preko 2 s), EER se smanjio u proseku za oko tri puta.

EER vrednosti prikazane na slikama 33-36. pokazuju da uređaji koji se koriste za snimanje glasa značajno utiču na performanse sistema za prepoznavanje zvučnika. U slučaju ovog eksperimenta, u laboratorijskim uslovima, najbolji rezultati su mereni integrisanim mikrofonom (laptop); nešto gore sa upotrebom 3G mobilnog telefona; a najgore sa jeftinim eksternim mikrofonom povezanim sa stonim računarom. Na primer, kombinacija ALIZE i integrisanog mikrofona rezultira 1,5 puta nižim EER -om za sve konfiguracije neusklađenosti pomoću spoljnog mikrofona.

8. ZAKLJUČAK

Prepoznavanje govornika privlači veliko interesovanje poslednju deceniju. Intenzivan rast novih tehnologija za prepoznavanje govora i govornika nam govori da je još uvek dug istraživački put pred nama. Osnovni cilj istraživanja je bio nalaženje odgovora na pitanje koliko uobičajne varijacije u interakciji korisnika sa sistemom za prepoznavanje korisnika na osnovu glasa mogu uticati na degradaciju funkcionalnih performansi takvih sistema, kao gradivnih elemenata menadžmenta identiteta u mrežnim aplikacijama u realnom vremenu. Ispitivanja su se vršila na nekoliko dostupnih rešenja sa otvorenim programskim kodom za prepoznavanje osoba na osnovu glasa, a koja implementiraju iste ili različite obradne paradigme za utvrđivanje autentičnosti izvora zvučnog zapisa.

Ovo istraživanje imalo je za cilj da utvrdi kako na performanse sistema za verifikaciju identiteta i kontrolu pristupa, zasnovanu na tehnikama prepoznavanja govornika, utiču izabrane tehnologije otvorenog koda, različite dužine izgovorenih rečenica, vrste uređaja za snimanje i neusklađenost između obuke i testnog jezika. Ova studija je ispitala performanse četiri popularna rešenja sistema za prepoznavanje govornika koji implementiraju različite pristupe i tehnike rešavanju problema. Nakon analize rezultata sistema prepoznavanja zvučnika, dobijenih u 144 različite testirane konfiguracije, došlo se do zaključka da je najuspešnije rešenje za prepoznavanje govornika otvorenog koda ALIZE. ALIZE alatka otvorenog koda za prepoznavanje govornika zasnovana je na Gaussovom modelu mešavine sa univerzalnim pozadinskim modelom (GMM-UBM). Podržava tehnike sa i-vektorskim modeliranjem i probabilističkom linearnom diskriminativnom analizom (PLDA).

Činjenica da su rezultati značajno bolji kada se engleski koristio i kao jezik za obuku i za testiranje sugeriše da su razmatrana rešenja optimizovana za upotrebu engleskog jezika. Ako se testirani softver otvorenog koda koristi na drugom jeziku, optimalne vrednosti konfiguracijskih parametara u softverskom paketu moraju biti pažljivo ispitane pre nego što se upotrebi. To implicira da performanse metoda prepoznavanja govornika zavise od jezika.

Analiza rezultata pokazuje da su vrednosti parametra EER u prvom približavanju obrnuto proporcionalne dužini rečenica. Dodatna istraživanja trebala bi ukazati na optimalnu dužinu rečenice pri kojoj se pristupa minimalnom mogućem EER -u.

Testirana softverska rešenja su veoma osetljiva na karakteristike upisa putem mikrofona na bilo kojem uređaju, što u praksi može značajno uticati na performanse sistema za prepoznavanje ili verifikaciju govornika. Poželjno je koristiti rešenje koje u predprocesiranju minimizira uticaj varijacija u karakteristikama uređaja za snimanje govora. Osim toga, rezultati predstavljeni u ovom radu ukazuju na moguće probleme koji se mogu pojaviti pri lokalizaciji aplikacije za chatbot koja zahteva verifikaciju govornika.

U daljem istraživanju bilo bi zanimljivo ispitati performanse sistema za prepoznavanje govornika koji koriste glas kao jedan od biometrijskih modaliteta, u dvofaktorskoj autentifikaciji (2FA) ili višefaktorskoj autentifikaciji (MFA).

8.1. OSTVARENI NAUČNI I STRUČNI DOPRINOSI

Rezultati istraživačkog procesa u predmetnoj oblasti, kao i rad na razvoju novog postupka, omogućili su više naučnih i stručnih doprinosa. Prvenstveno tokom istraživačkog procesa urađen je celovit pregled istraživačke oblasti prepoznavanja osoba na osnovu glasa, te ispitivanje identifikacije osobe na osnovu glasa putem četiri programa otvorenog koda, uz njihovu kritičku analizu. Na osnovu urađene kritičke analize identifikovan je prostor za nove naučne metode u smislu kakav je uticaj varijacija u interakciji korisnika na performanse sistema za prepoznavanje govornika.

U okviru predloženih metoda primenjene su nove linije kodova u četiri programa otvorenog koda u postupku akvizicije podataka u govoru. Za potrebe transformacije prikupljenih podataka govornika obrađeni su kroz tri modela uređaja, kroz dva jezika i kroz veličinu vremenskog uzorka za uticaj na krajnju preciznost u prepoznavanju.

Radi emprijske provere valjanosti formulisanog postupka kreirana je zasebna multimedijalna baza biometrijskih podataka. Na osnovu toga u kombinaciji sa evaluacijom istraženi su raznoliki segmenti kako funkcioniše program. Istraživanjem se došlo do rezultata ako je uređaj na treningu i testu isti, preciznost sistema je veća. Isti slučaj je i sa jezikom. Dobijeni rezultati su ukazali da što je kraće vreme govornog uzorka, postiže se efikasnije poboljšanje preciznosti sistema.

Na osnovu empirijske evaluacije predloženog postupka može se zaključiti da postupak ima puno potencijala, da ima osnovu za primenu u praksi i da postiže zadovoljavajuće stope prepoznavanja.

8.2. MOGUĆNOSTI PRIMENE

Praktična upotreba je odraz efikasnsoti biometrijskog sistema. Primena u stvarnom životu diriguje ocenu kvaliteta.

Urađen rad na temu uticaj varijacija u interakciji korisnika na performanse sistema za prepoznavanje govornika je doprineo tome da se može i besplatno poboljšati opšta sigurnost nekog sistema, jer se koriste programi otvorenog koda.

Može da se primeni u oblasti bankarstva, a jedan od njih je i bankomat, jer se PIN vrlo lako može pogrešiti ili zaboraviti, dok glas kao biometrijska karakteristika čoveka je individualna. Takođe, primenjivo je u sigurnosti transakcija novca putem kreditnih kartica, kao i za proveru identiteta na daljinu (kada se ne nalazimo u zemlji). Australijska vlada već dvadesetak godina koristi prepoznavanje govornika za autentifikaciju koristeći telefonski razgovor (Summerfield, Dunstone, & Summerfield, 2008). Svakako se ostavlja prostor za istraživanje u kombinaciji sa drugom biometrijskom tehnikom.

Bezbednosne agencije mogu da koriste jedan takav sistem putem prisluškivanja, te se samim tim prepozna određena osoba od značaja.

Multinacionalne kompanije često održavaju konferencijske razgovore putem grupnog online telefonskog razgovora, gde se susreću ljudi iz celog sveta koji se međusobno ne poznaju. U tom razgovoru postoji mogućnost prepoznavanja govornika.

Jedna od glavnih prednosti prepoznavanja govornika u odnosu na druge biometrijske tehnike je niska cena, visoka prihvatljivost i neinvazivna tehnika uzimanja uzoraka.

8.3. MOGUĆI DALJI PRAVCI ISTRAŽIVANJA

Navedeni primer obrade podataka za prepoznavanje govornika je pokazao visok nivo u razvojnom okruženju, te se kao takav pokazao kao valjan primer za praktičnu upotrebu. Dobijeni rezultati se smatraju korisnim za dalja unapređenja samog sistema za prepoznavanje govornika, gde se stvara prostor za dopunu novih funkcionalnosti.

Potrebno je raditi na sofisticiranosti same tehnologije, kao i na njenoj integraciji sa realnim životom. Kao što smo videli u ovom radu, moguće je dobro protumačiti govor, ali dalji korak bi bio kreirati razgovorni intefejs, što je svakako moguće u open-source tehnologijama.

Sistem za prepoznavanje govornika i dalje ima nedostatke koji se mogu dodatno smanjiti sprovođenjem istraživanja na polju poddomenskih tehnika, kao i spajanjem sa drugim biometrijskim sistemima kako bi se postigla bolja pouzdanost u autentičnosti. Svakako da je glavna aplikacija rešena i u ovom radu, a ona je u oblasti kontrole pristupa, gde se od govornika zahteva da nešto kaže i samim tim mu se potrvdi identitet, kako bi mu se potvrdio pristup određenim informacijama, kontroli sistema, nekim ograničenim servisima u različitim domenima koji zahtevaju bezbednost.

Za potrebe daljeg istraživanja i uanpređenje procesa potrebno je proširiti multimedijalnu biometrijsku bazu podataka koja će sadržati veći broj korisnika.

9. LITERATURA

1. A. Anjos, L. E.-C. (2012). *Bob: a free signal processing and machine learning toolbox for researchers.* ACM International Conference on Multimedia.
2. Achillas, C. B. (2019). *Voice-driven fleet management system for agricultural operations.* Information Processing in Agriculture.
3. Alam, J., Bhattacharya, G., & Kenny, P. (2018). Speaker Verification in Mismatched Conditions with Frustratingly Easy Domain Adaptation. *In Proceedings of the Odyssey 2018 The Speaker and Language Recognition Workshop.* France: Les Sables d'Olonne.
4. Alexakis, G. P. (2019). *Control of Smart Home Operations Using Natural Language Processing, Voice Recognition and IoT Technologies in a Multi-Tier Architecture.* Designs, 3(3), 32.
5. Alluri, K., & Vuppala, A. (2020). A study on the emotional state of a speaker in voice bio-metrics. *In Advances in Ubiquitous Computing*, 223-236.
6. Arnold, R. (1991). *Betriebliche Weiterbildung.* Bad Heilbrunn/Obb.: Klinkhardt.
7. Asghar Taheri, M. R. (February 2005). *Fuzzy Hidden Markov Models for speech recognition on based FEM Algorithm.* Transaction on engineering Computing and Technology V4, IISN,1305-5313.
8. Babaei, H., Molalapata, O., & Pandor, A. (2012). Face Recognition Application for Automatic Teller Machines (ATM). *ICIKM 45,* (str. 211-216).
9. Bargi, A., Da Xu, R., & Piccardi, M. (2017). AdOn HDP-HMM: An adaptive online model for segmentation and classification of sequential data. *IEEE Transactions on Neural Networks and Learning Systems, 29(9)* (str. 3953-3968). IEEE.
10. Beigi, H. (2011). Speaker recognition. In Fundamentals of Speaker Recognition . U H. Beigi, *Speaker recognition* (str. 543-559). Boston, MA.: Springer.

11. Bell, P., Fainberg, J., Klejch, O., Li, J., Renals, S., & Swietojanski, P. (2020). Adaptation algorithms for neural network-based speech recognition: An overview. *IEEE Open Journal of Signal Processing, 2*, 33-66.

12. Benson, B. (2018). Fingerprint Not Recognized: Why the United States Needs to Protect Biometric Privacy. *North Carolina Journal of Law & Technology, 19(4)*, 161.

13. Bolle, R., Ratha, N., & Pankanti, S. (2004). Performance evaluation in 1: 1 biometric engines. *In Chinese Conference on Biometric Recognition* (str. 27-46). Berlin, Heidelberg: Springer.

14. Buzo, A. e. (2014). *An Automatic Speech Recognition solution with speaker identification support.* Communications (COMM), 2014 10th International Conference on. IEEE.

15. Cai, D., Cai, W., & Li, M. (2020). Within-sample variability-invariant loss for robust speaker recognition under noisy environments. *In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (str. 6469-6473). IEEE.

16. Cai, D., Qin, X., & Li, M. (2019). Multi-Channel Training for End-to-End Speaker Recognition Under Reverberant and Noisy Environment. *In Interspeech*, 4365-4369.

17. Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in cognitive sciences, 11(12)*, 535-543.

18. Cao, Y., & Yang, L. (2010). A survey of identity management technology. *In 2010 IEEE International Conference on Information Theory and Information Security* (str. 287-293). IEEE.

19. Cavoukian, A. (1999). Consumer biometric applications: A discussion paper. *THE Office*.

20. Cernak, M., Komaty, A., Mohammadi, A., Anjos, A., & Marcel, S. (n.d.). Bob speaks kaldi. *Proc. of Interspeech*, (str. 2017).

21. Cowger, J. (2020). Friction ridge skin: comparison and identification of fingerprints. *CRC Press*.

22. Curatelli, F., Mayora-Ibarra, O., & Carotenuto, D. (1999). SPEAR, A Modular Tool for Speech Signal Processing and Recognition. *Proceeding of WISP'99.*

23. Dash, D., Wisler, A., Ferrari, P., & Wang, J. (2019). Towards a Speaker Independent Speech-BCI Using Speaker Adaptation. *In INTERSPEECH*, 864-868.

24. Ding, S., Wang, Q., Chang, S., Wan, L., & Moreno, I. (2019). Personal VAD: Speaker-conditioned voice activity detection. *arXiv preprint arXiv:1908.04284.*

25. Dunphy, P. &. (2018). *A first look at identity management schemes on the blockchain.* IEEE Security & Privacy, 16 (4), 20-29.

26. E. Khoury, L. E. (2013). *Bi-modal biometric authentication on mobile phones in challenging conditions.* Image and Vision Computing.

27. El-Moneim, S., Sedik, A., Nassar, M., El-Fishawy, A., Sharshar, A., Hassan, S., & Elabyad, G. (2021). Text-dependent and text-independent speaker recognition of reverberant speech based on CNN. *International Journal of Speech Technology, 24(4),* 993-1006.

28. Faúndez-Zanuy, M., & Rodríguez-Porcheron, D. (2022). Speaker recognition using residual signal of linear and nonlinear prediction models. *arXiv preprint arXiv:2203.09231*.

29. Femila, M., & Irudhayaraj, A. (2011). Biometric system. *In 2011 3rd International Conference on Electronics Computer Technology Vol. 1*, 152-156.

30. Fine, S., Singer, Y., & Tishby, N. (1998). The hierarchical hidden Markov model: Analysis and applications. *Machine learning, 32(1),* 41-62.

31. Furui, S. (1981). Cepstral analysis technique for automatic speaker verification. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 29(2)*, 254-272.

32. Gada, J., Rao, P., & Samudravijaya, K. (2013). Confidence measures for detecting speech recognition errors. *In 2013 National Conference on Communications (NCC* (str. 1-5). IEEE.

33. Gaikwad, S., Gawali, B., & Yannawar, P. (2010). A review on speech recognition technique. *International Journal of Computer Applications, 10(3),* 16-24.

34. Gao, J., & Xiang, L. (2010). Local preserving projections andwithin-class scatter based semi-supervised support vector machines. *In 2010 3rd International Conference on Computer Science and Information Technology* (str. 267-270). IEEE.

35. Gao, Z. X. (2018, June). *Blockchain-based identity management with mobile device. In Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems (pp. 66-70). .* ACM.

36. Gin-der Wu, Y. L. (2008). *A Register Array based Low power FFT Processor for speech recognition.* Taiwan: Department of Electrical engineering national Chi Nan university Puli ,545 .

37. Giotis, A. P. (2017). *A survey of document image word spotting techniques.* Pattern Recognition, 68, 310-332.

38. Grozdić, Đ. (2017). *Primena neuralnih mreža u prepoznavanju šapata, doktorska disertacija.* Preuzeto sa Phaidra BG: https://phaidrabg.bg.ac.rs/open/o:17276

39. Hanna, K. J. (2018). *Systems and methods for an incremental, reversible and decentralized biometric identity management system. .* U.S. Patent Application No. 10/078,758.

40. Hansen, J., & Hasan, T. (2015). Speaker recognition by machines and humans: A tutorial. *IEEE Signal processing magazine, vol. 32, no. 6,* 74–99.

41. Hermansky, H., Morgan, N., Bayya, A., & Kohn, P. (1991). RASTA-PLP speech analysis. *In Proc. IEEE Int'l Conf. Acoustics, speech and signal processing Vol. 1* (str. 121-124). IEEE.

42. Hofbauer, H., & Uhl, A. (2016). Calculating a boundary for the significance from the equal-error rate. *In 2016 International Conference on Biometrics (ICB)* (str. 1-4). IEEE.

43. Hoffmann, R. (2010). On the development of early vocoders. *In 2010 Second Region 8 IEEE Conference on the History of Communications* (str. 1-6). IEEE.

44. Huang, X. &. (1993). *On speaker-independent, speaker-dependent, and speaker-adaptive speech recognition.* IEEE Transactions on Speech and Audio processing, 1(2), 150-157.

45. ISO/IEC, 2.-1. (2011). *Security Techniques—A Framework for Identity Management—Part 1: Terminology and Concepts,.* ISO/IEC.

46. Jacob, I., Betty, P., Darney, P., Raja, S., Robinson, Y., & Julie, E. (2021). Biometric template security using DNA codec based transformation. *Multimedia Tools and Applications, 80(5),* 7547-7566.

47. Jadhav, A., & Dharwadkar, N. (2018). A Speaker Recognition System Using Gaussian Mixture Model, EM Algorithm and K-Means Clustering. *International Journal of Modern Education & Computer Science, 10(11)*.

48. Jain, A. K. (1999). *Biometrics: Personal Identification.* Dordrecht: Kluwer Academic Publishers.

49. Jain, A., Hong, L., & Pankanti, S. (2000). Biometric identification. *Communications of the ACM 43 (2),* 90-98.

50. Javed, I., Alharbi, F., Bellaj, B., Margaria, T., Crespi, N., & Qureshi, K. (2021). Health-ID: a blockchain-based decentralized identity management for remote healthcare. *In Healthcare (Vol. 9, No. 6),* 712.

51. Jokić, I. (2009). *Problematika automatskog prepoznavanja govornika.* Preuzeto sa Infoteh: https://infoteh.etf.ues.rs.ba/zbornik/2009/radovi/B-III/B-III-11.doc

52. Kalinli, O., Bhattacharya, G., & Weng, C. (2019). Parametric Cepstral Mean Normalization for Robust Speech Recognition. *In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP* (str. 6735-6739). IEEE.

53. Kamath, U., Liu, J., & Whitaker, J. (2019). *Deep learning for NLP and speech recognition (Vol. 84).* Switzerland: Springer.

54. Kambeyanda, D., Singer, L., & Cronk, S. (1997). Potential problems associated with use of speech recognition products. *Assistive Technology, 9(2),* 95-101.

55. Kaushal, N., & Kaushal, P. (2011). Human identification and fingerprints: a review. *J Biomet Biostat, 2(123),* , 2.

56. Kenny, P. a. (2004). *Disentangling speaker and channel effects in speaker verification.* IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP, pp. 37–40.

57. Keo, S., Kak, S., Shiga, Y., Kato, H., & Kawai, H. (2019). HMM-based TTS System Framework. *In 2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)* (str. 1-1). IEEE.

58. Khoury, E. L. (2014). *Acoustics, Speech and Signal Processing (ICASSP).* IEEE International Conference.

59. Kinnunen, T. a. (2010). *An overview of text-independent speaker recognition: From features to supervectors.* Speech communication 52.1 12-40.

60. Kinnunen, T., & Li, H. (2010). An overview of text-independent speaker recognition: From features to supervectors. *Speech communication, 52(1),* 12-40.

61. Kleinberg, K., Vanezis, P., & Burton, A. (2007). Failure of anthropometry as a facial identification technique using high-quality photographs. *Journal of forensic sciences, 52(4),*, 779-783.

62. Kolak, A. (2009). *Sustav za automatsko prepoznavanja izgovora matičnog broja studenta.* Zagreb: Sveučilište u Zagrebu, Fakultet elektronike i računarstva.

63. Kolak, A. (2009). *SUSTAV ZA AUTOMATSKO PREPOZNAVANJE IZGOVORA MATIČNOG BROJA STUDENATA.* Preuzeto sa BIB: https://www.bib.irb.hr/451087/download/451087.Zavrsni_Kolak_Antonio_0036427170.pdf

64. Kong, Y., Somarowthu, A., & Ding, N. (2015). Effects of spectral degradation on attentional modulation of cortical auditory responses to continuous speech. *Journal of the Association for Research in Otolaryngology, 16(6)*, 783-796.

65. Kovačević, M. (2012). *Simulacija daljinskog upravljača sa podrškom za prepoznavanje i mogućnost upravljanja glasom na Android platformi (završni rad).* Preuzeto sa RTK-RK: https://www.rt-rk.uns.ac.rs/sites/default/files/e12775_milan_kovacevic.pdf

66. Kuršumlija, G. (2012). *Obrada zvuka i zvučni (audio) formati.* Preuzeto sa Gimnazija Kuršumlija: https://gimnazijakursumlija.files.wordpress.com/2012/03/obrada-zvuka-i-zvuc48dni-audio-formati.doc

67. Larcher, A. e. (2013). *ALIZE 3.0-open source toolkit for state-of-the-art speaker recognition.* INTERSPEECH.

68. Latinović, O. (2017). SPEAKER RECOGNITION FOR THE INTERNET OF THINGS AS SECURITY MEASURE. *XL VI Simpozijum o operacionim istraživanjima* (str. 165-169). Zlatibor: Visoka građevinsko-geodetska škola, Beograd. Preuzeto sa http://symopis.vggs.rs/razno/Zbornik_radova_SYM-OP-IS_2017.pdf

69. Lattore, J., Iwano, K., & Furui, S. (2005). Polyglot synthesis using a mixture of monolingual corpora. *In Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing* (str. 1). IEEE.

70. Liu, F., Tong, J., Mao, J., Bohn, R., Messina, J., Badger, L., & Leaf, D. (2011). NIST cloud computing reference architecture. *NIST special publication*, str. 1-28.

71. Ma, B., Meng, H., & Mak, M. (2007). Effects of device mismatch, language mismatch and environmental mismatch on speaker verification. *In 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, vol.4* (str. IV-301). IEEE.

72. Maghsoudi, J., & Tappert, C. (2016). A Behavioral Biometrics User Authentication Study Using Motion Data from Android Smartphones. *2016 European Intelligence and Security Informatics Conference (EISIC)* (str. 184-187). IEEE.

73. Malik, J., Girdhar, D., Dahiya, R., & Sainarayanan, G. (2014). Reference threshold calculation for biometric authentication. *International Journal of Image, Graphics and Signal Processing, 6(2)*, 46.

74. Malik, M., Malik, M., Mehmood, K., & Makhdoom, I. (2021). Automatic speech recognition: a survey. *Multimedia Tools and Applications, 80(6)*, 9411-9457.

75. Markowitz, J. (1999). *Introduction to SVAPI.* SVAPI Working Group.

76. Markowitz, J. (2008). Myths and Misunderstandings about Speaker Authentication.

77. Matanović, A. (2018). OSNOVE KRIPTOVALUTA I BLOKČEIN TEHNOLOGIJE. *portal mojafirma.rs*.

78. McLaren, M., Lei, Y., & Ferrer, L. (2015). Advances in deep neural network approaches to speaker recognition. *In 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (str. 4814-4818). IEEE.

79. Mesarović, S. (2007). *Multimodalni biometrijski sistemi, magistarska teza.* Preuzeto sa Singipedia: https://singipedia.singidunum.ac.rs/preuzmi/42151-multimodalni-biometrijski-sistemi/2233

80. Miguel, A., Lleida, E., Rose, R., Buera, L., Saz, O., & Ortega, A. (2008). Capturing local variability for speaker normalization in speech recognition. *IEEE transactions on audio, speech, and language processing, 16(3)* (str. 578-593). IEEE.

81. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems, 26*.

82. Milošević, M. (2020). *Identifikacija govora u uslovima emotivnog govora, doktorska disertacija.* Preuzeto sa Fedora: https://fedorabg.bg.ac.rs/fedora/get/o:23058/bdef:Content/download

83. Milovanović, M. (2013). *PRIMENA CBIR TEHNIKA U BIOMETRIJSKOJ, doktorska disertacija.* Preuzeto sa Fedora: https://fedorabg.bg.ac.rs>bdef:Content>download

84. Mokhov, S. (2006). *On Design and Implementation of Distributed Modular Audio Recognition Framework: Requirements and Specification Design Document.* Department of Computer Science and Software Engineering, Concordia University.

85. Mondal, S., Bours, P., & Idrus, S. (2013). Complexity measurement of a password for keystroke dynamics: Preliminary study. *In Proceedings of the 6th International Conference on Security of Information and Networks* , 301-305.

86. Monroe, D. (2012). *Biometrics Metrics Report v3. 0.*

87. Morales, N. J. (2005). *MFCC Compensation for Improved Recognition of Filtered and Band-Limited Speech.* ICASSP (1).

88. Najka, R. (2018). An overview of automatic speaker verification system. *Intelligent Computing and Information and Communication*, 603-610.

89. Nikias, C., & Mendel, J. (1993). Signal processing with higher-order spectra. *IEEE Signal processing magazine, 10(3)*, 10-37.

90. Nimac, L. (2007). *Pregled biometrijskih metoda autentifikacije.* Preuzeto sa Scribd: https://www.scribd.com/doc/50946883/Pregled-biometrijskih-metoda-autentifikacije

91. Niwatkar, A., & Kanse, Y. (2020). Feature Extraction using Wavelet Transform and Euclidean Distance for speaker recognition system. *In 2020 International Conference on Industry 4.0 Technology (I4Tech)* (str. 145-147). IEEE.

92. Ostapenko, A., Wintner, S., Fricke, M., & Tsvetkov, Y. (2022). *Speaker Information Can Guide Models to Better Inductive Biases: A Case Study On Predicting Code-Switching.* arXiv preprint arXiv:2203.08979.

93. Park, A., & Hazen, T. (2002). ASR dependent techniques for speaker recognition. *In International conference on Spoken Language Processing*, (str. 1337-1340).

94. Paunović, S., Starčević, D., & Nešić, L. (2013). Identity management and witness protection system. *Management, 66*, 27-37.

95. R.Klevansand, R. (1997). *Voice Recognition.* Boston, London : Artech House.

96. R.Klevansand, R. (1997). *Voice Recognition.* Boston, London: Artech House.

97. Rabiner, L., & Juang, B. (1993). *Fundamentals of speech recognition.* Prentice-Hall, Inc..

98. Radmanić, A. (2019). *Fotografija u muzejskoj dokumentaciji, master rad.* Preuzeto sa Repozitorij: https://repozitorij.unizg.hr/islandora/object/ffzg%3A577/datastream/PDF/view

99. Rafieee, M., & Khazaei, A. (2010). A novel model characteristics for noise-robust automatic speech recognition based on HMM. *In 2010 IEEE International Conference on Wireless Communications, Networking and Information Security* (str. 215-218). IEEE.

100. Rane, S. (2014). Standardization of biometric template protection. *IEEE MultiMedia, 21(4),* 94-99.

101. Ranjan, R., & Dubey, R. (2016). Isolated word recognition using HMM for Maithili dialect. *In 2016 international conference on signal processing and communication (ICSC)* (str. 323-327). IEEE.

102. Reid, P. (2004). *Biometrics for network security.* Prentice Hall Professional.

103. Reynolds, D. (1995). Automatic speaker recognition using Gaussian mixture speaker models. *The Lincoln Laboratory Journal*.

104. Reynolds, D. (2002). An overview of automatic speaker recognition technology. *In 2002 IEEE international conference on acoustics, speech, and signal processing Vol. 4* (str. IV-4072). IEEE.

105. Sadjadi, O., Greenberg, C., Singer, E., Mason, L., & Reynolds, D. (2021, Decembar 20). *Speaker Recognition Evaluation Plan*. Preuzeto sa NIST SRE: https://sre.nist.gov/sre21

106. Sadjadi, S., Kheyrkhah, T., Tong, A., Greenberg, C., Reynolds, D., Singer, E., . . . Hernandez-Cordero, J. (2017). The 2016 nist speaker recognition evaluation. *Interspeech*, 1353–1357.

107. Sadoki Furuki, T. I. (2005). *Cluster-based Modeling for Ubiquitous Speech Recognition.* Department of Computer Science Tokyo Institute of Technology Interspeech .

108. Saquib, Z., Salam, N., Nair, R., Pandey, N., & Joshi, A. (2010). A survey on automatic speaker recognition systems. *Signal Processing and Multimedia*, 134-145.

109. Simpson, R. (2013). *Computer Access for People with Disabilities: A Human Factors Approach.* CRC Press.

110. Song, Z., Mokhov, S., Song, M., & Mudur, S. (2018). Creative use of signal processing and MARF in ISSv2 and beyond. *In ACM SIGGRAPH 2018 Posters*, 1-2.

111. Souza, J., Medeiros, A., Holanda, G., Rego, P., & Reboucas Filho, P. (2021). Fingerprint Classification Based on the Henry System via ResNet. *In International Conference on Systems, Signals and Image Processing* (str. 15-28). Springer, Cham.

112. Steve Young, G. E. (2001). *HTK Book.* Cambridge University Engineering Department.

113. Summerfield, R., Dunstone, T., & Summerfield, C. (2008). Speaker verification in a multi-vendor environment. *In W3C workshop on speaker identification and verificatio.* SIV.

114. Sztahó, D., Szaszák, G., & Beke, A. (2019). Deep learning methods in speaker recognition: a review. *arXiv preprint arXiv:1911.06615.*

115. Tait, B. (,2021). Aspects of Biometric Security in Internet of Things Devices. *In Digital Forensic Investigation of Internet of Things (IoT) Devices* , 169-186.

116. Togneri, R., & Pullella, D. (2011). An overview of speaker identification: Accuracy and robustness issues. *IEEE circuits and systems magazine, 11(2)*, 23-61.

117. Tur, G., & De Mori, R. (2011). *Spoken language understanding: Systems for extracting semantic information from speech.* John Wiley & Sons.

118. Varchol, P., & Levicky, D. (2007). Using of hand geometry in biometric security systems. *Radioengineering-Prague-, 16(4)*, 82.

119. Vaseghi, S. (1996). Spectral subtraction. *In Advanced Signal Processing and Digital Noise Reduction*, 242-260.

120. Vaseghi, S. (2008). *Advanced digital signal processing and noise reduction.* John Wiley & Sons.

121. Venkateswarlu, R., Raviteja, R., & Rajeev, R. (2012). The performance evaluation of speech recognition by comparative approach. U R. Venkateswarlu, R. Raviteja, & R. Rajeev, *Advances in Data Mining Knowledge Discovery and Applications.* IntechOpen.

122. Wang, J., Ge, W., Snow, D., Mitra, P., & Giles, C. (2010). Determining the sexual identities of prehistoric cave artists using digitized handprints: a machine learning approach. *In Proceedings of the 18th ACM international conference on Multimedia* , (str. 1325-1332).

123. Wenndt, S., & Shamsunder, S. (1997). Bispectrum features for robust speaker identification. *In 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing Vol. 2* (str. 1095-1098). IEEE.

124. Wolf, J. (1972). *Efficient acoustic parameters for speaker recognition.* Journal of the Acoustic Society of America.

125. Xiong, F., Barker, J., & Christensen, H. (2019). Phonetic analysis of dysarthric speech tempo and applications to robust personalised dysarthric speech recognition. *In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (str. 5836-5840). IEEE.

126. Xu, X., Weber, I., & Staples, M. (2017). *Architecture for blockchain applications.* Heidelberg: Springer.

127. You, S., Lu, H., & Shixing, L. (2009). Research on rule definition and engine for general text processing. *In 2009 4th International Conference on Computer Science & Education* (str. 1095-1100). IEEE.

128. Yu, K., Mason, J., & Oglesby, J. (1995). Speaker recognition using hidden Markov models, dynamic time warping and vector quantisation. *IEE Proceedings-Vision, Image and Signal Processing, 142(5)*, (str. 313-318).

129. Zheng, T., & Li, L. (2017). *Robustness-related issues in speaker recognition.* Springer.

LISTA TABELA

LISTA SLIKA

BIOGRAFIJA

Olja Latinović rođena je 24. aprila 1988. godine u Prijedoru. Nakon završene gimnazije u Prijedoru (2006. godine) upisuje redovne studije na Panevropskom univerzitetu „Apeiron“ u Banja Luci – smer Inženjering informacionih tehnologija. Odličnom ocenom na odbrani diplomskog rada „MS SQL baza bibliotečkog informacionog sistema sa interfejsom u C#“, te 2010. postaje diplomirani inženjer informacionih tehnologija. Nakon završetka osnovnih studija u Banja Luci, upisuje na Fakultetu organizacionih nauka Univerziteta u Beogradu diplomske akademske - master studije, studijski program Informacioni sistemi i tehnologije, i sa ocenom deset odbranom master rada stiče zvanje mastera. Doktorske studije na Fakultetu organizacionih nauka upisuje, studijski program Informacioni sistemi i kvantitativni menadžment, izborno područje Informacioni sistemi upisala je 2012.g.

Od 2009. do 2013. godine učestvuje u nastavi kao stručni saradnik za predmete Baze podataka, Projektovanje informacionih sistema, Poslovne aplikacije, Informacione tehnologije, te uskoro postaje asistent, zatim i viši asistent na Panevropskom univerzitetu „Apeiron“. Pored toga, učestvuje u organizaciji međunarodnog naučno-stručnog skupa Informacione tehnologije u elektronskom obrazovanju (ITeO) koji se svake godine održava u Banja Luci, te u radu seminara „Oblici i metode interaktivne nastave sa studentima“. 2013. godine počinje da radi u softverskoj firmi „Breza software engineering“ u Beogradu, kao softverski inženjer, gde je angažovana na realizaciji više projekata. U međuvremenu objavljuje naučne radove, najviše u sferi biometrije, sa saradnicima, asistentima i profesorima iz multimedijalne laboratorije sa Fakulteta organizacionih nauka. Pored toga, stiče zvanične „Oracle“ sertifikate kao Oracle Database 11g Performance Tuning Certified Expert, te Oracle Database 11g Administrator Certified Professional, i Oracle Certified Professional Java SE 7 Programmer. Od 2017. godine je radno angažovana na Univerzitetu „Union – Nikola Tesla“ u Beogradu kao asistent na informatičkoj grupi predmeta.

Govori srpski, engleski, nemački i francuski jezik.

Za vreme diplomskih i doktorskih akademskih studija učestvovala je kao student saradnik u radu Laboratorije za multimedijalne komunikacije na projektu “Primena multimodalne biometrije u menadžmentu identiteta”, TR32013, Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije.

Изјава о ауторству

Име и презиме аутора: **Оља (Бранко) Крчадинац**

Број индекса: **5020/2012**

Изјављујем

да је докторска дисертација под насловом

„Утицај варијација у интеракцији корисника на перформансе система за препознавање говорника“

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршила ауторска права и користила интелектуалну својину других лица

Потпис аутора

У Београду, _____

_____

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора:	**Оља (Бранко) Крчадинац**
Број индекса:	**5020/2012**
Студијски програм:	**Информациони системи и менаџмент**
Наслов рада:	**Утицај варијација у интеракцији корисника на перформансе система за препознавање говорника**
Ментор:	**др Душан Старчевић, професор емеритус**

Ја, Оља Крчадинац

Изјављујем да је штампана верзија мог докторског рада истоверна електронској верзији коју сам предала ради похрањивања у **Дигиталном репозиторијуму Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

У Београду, _____

_____

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

„Утицај варијација у интеракцији корисника на перформансе система за препознавање говорника“

која је моје ауторско дело.

Дисертацију са свим прилозима предала сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио.

1. Ауторство (CC BY)

2. Ауторство – некомерцијално (CC BY-NC)

3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)

4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)

5. Ауторство – без прерада (CC BY-ND)

6. Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених лиценци. Кратак опис лиценци је саставни део ове изјаве).

Потпис аутора

У Београду, _____

_____